

Matched Filtering from Limited Frequency Samples

Armin Eftekhari,^c Justin Romberg,^g and Michael B. Wakin^{c*}

^c Dept. of Electrical Engineering and Computer Science, Colorado School of Mines

^g School of Electrical and Computer Engineering, Georgia Tech

January 2011; revised July 2012

Abstract

In this paper, we study a simple correlation-based strategy for estimating the unknown delay and amplitude of a signal based on a small number of noisy, randomly chosen frequency-domain samples. We model the output of this “compressive matched filter” as a random process whose mean equals the scaled, shifted autocorrelation function of the template signal. Using tools from the theory of empirical processes, we prove that the expected maximum deviation of this process from its mean decreases sharply as the number of measurements increases, and we also derive a probabilistic tail bound on the maximum deviation. Putting all of this together, we bound the minimum number of measurements required to guarantee that the empirical maximum of this random process occurs sufficiently close to the true peak of its mean function. We conclude that for broad classes of signals, this compressive matched filter will successfully estimate the unknown delay (with high probability, and within a prescribed tolerance) using a number of random frequency-domain samples that scales inversely with the signal-to-noise ratio and only logarithmically in the observation bandwidth and the possible range of delays.

1 Introduction

1.1 Random Sampling and Compressive Signal Processing

Over the last few decades, the development of cheap, flexible, and powerful digital signal processing (DSP) architectures has enabled the acquisition and analysis of increasingly rich data sets. One of the key principles behind the DSP revolution is the fundamental work by Nyquist, Whittaker, and Shannon in characterizing the minimum number of discrete-time samples required to fully capture the information in a bandlimited continuous-time signal. Unfortunately, many real-world signals of interest may have very high bandwidth, which can severely complicate the practical task of sampling a signal at its Nyquist rate [21, 35].

*Email: aeftekha@mines.edu, jrom@ece.gatech.edu, mwakin@mines.edu. This work was partially supported by NSF Grants CCF-0830320 and CNS-0910592 and by DARPA grant FA8650-08-C-7853. Submitted to the *IEEE Transactions on Information Theory* on January 13, 2011.

The recently developed theory of Compressive Sensing (CS) [5, 9] suggests that if a signal is structured, then we can acquire it by taking samples well below its Nyquist rate. CS relies on two fundamental principles: first, that many signals have much lower complexity than is suggested by their bandwidth (typically this is embodied in a sparse representation for the signal within some basis), and second, that such signals may safely be sampled below their Nyquist rate if the traditional uniform time-domain sampling procedure is replaced with a generalized linear measurement operator (typically this operator contains some degree of randomness).

The CS theory has benefited from several powerful and elegant tools for probabilistic analysis relating to the theory of empirical processes. The essential condition (the *restricted isometry property* [4]) that guarantees sparse recovery from observations through a random matrix can be recast as a bound on a random process — this formulation, first put forth in [34], is particularly useful when the compressive measurement system is structured [28, 30, 32, 33, 35]. In these works, the Dudley inequality [12], a classical tool which relates the supremum of a random process to the geometry of its index set, is used to bound the expected supremum of the process, and strong tail bounds are established that control the deviation from the average behavior. To date, almost all of the work along these lines has focused on providing guarantees for signal recovery from compressive measurements.

There are many applications, however, where we are not interested in a full-scale recovery of a signal. Instead, we may wish only to estimate some key parameters (or “features”) in order to solve an inference problem that does not demand full knowledge of the signal. It has been demonstrated that random measurements can again be very useful in such settings. Just as certain low-complexity signals can be fully recovered from random measurements, certain low-complexity questions can be answered about (possibly arbitrary) signals directly from random measurements without first recovering the signal. Some initial steps in this direction have been concerned with compressive detection, classification, estimation, and filtering [6, 7, 11, 16, 17]. Compared to alternative techniques that base their inference on a full set of Nyquist samples, compressive inference techniques can show slightly diminished accuracy because fewer statistics are measured concerning the signal. In exchange, the acquisition hardware can potentially be much simpler and consume less power. In addition, we maintain the ability to adapt to future information we may learn the problem at hand; from a single set of random measurements, a number of different inference problems may be solved concerning a number of possible candidate signals.

In this paper, we study the problem of matched filtering (i.e., estimating the unknown delay and amplitude of a known template signal) from the compressive viewpoint. In particular, we derive strong bounds on the performance of a compressive matched filter by bringing in some of the same probabilistic tools that have been so fruitful in CS recovery analysis. To do this, we show that the compressive matched filtering problem can be reduced to controlling the supremum of a certain random process, which we approach through a specialized version of the Dudley inequality.

1.2 Matched Filtering from Limited Frequency Samples

1.2.1 The Compressive Matched Filter

The problem that we consider is formally stated as follows. Suppose we make observations of the continuous-time signal $A \cdot s_0(t - \tau_0)$, where $s_0(t)$ is a known signal template, $\tau_0 \in T \subset \mathbb{R}$ is the unknown delay (the *time-of-arrival*), and $A \in \mathbb{R}$ or \mathbb{C} is the unknown amplitude. Given these observations, we would like to estimate τ_0 and A .

The optimal solution to this problem is given by the *matched filter*. All shifts of the known template $s_0(t)$ are correlated against the incoming signal, and the estimated time-of-arrival is the shift that yields the maximum correlation. The matched filter is typically implemented in one of two ways: either with a specialized analog circuit that performs the correlation and then detects the peaks, or by sampling the signal and calculating the correlation function digitally. The advantage of the digital approach is the flexibility it offers; we can perform matched filtering against many different waveforms from the same set of samples, in case $s_0(t)$ is not known in advance (but belongs to a collection of candidates). If the signal $s_0(t)$ is concentrated in time, the sampling rate must be commensurately high to accurately estimate τ_0 . Applications such as high-frequency radar or ultra-wideband communications can require sampling rates of hundred of millions, or even billions, of samples per second. Taking and processing samples at these kinds of rates is costly in terms of hardware complexity and power consumption.

Working from compressed samples gives us a more elegant solution. In this paper we analyze a simple correlation-based estimator that operates using a small number m of noisy samples of spectrum of the received signal, with locations that are drawn randomly from a uniform distribution on some interval Ω in the frequency domain. In one of our main results (Corollary 6), we prove that for broad classes of signals, this compressive matched filter will successfully estimate τ_0 (with high probability, and within a prescribed tolerance) using a number of random frequency-domain samples that scales inversely with the signal-to-noise ratio and only logarithmically in the observation bandwidth $|\Omega|$ and the possible range of delays $|T|$. Our results help validate the use of compressive measurements for capturing important signal information. This acquisition scheme also offers us flexibility in that it depends only on very broad characteristics of the signal; it is universally effective for all $s_0(t)$ which are spread out over the band Ω .

We note that the use of randomized measurements in the frequency domain is not unprecedented. One of the original motivating problems for CS, for example, came from magnetic resonance imaging (MRI), where the goal is to reconstruct an image from a partial set of Fourier coefficients [5, 25]. Randomized frequency-domain measurements are also standard in CS problems where the signal to be recovered is sparse in the time domain [5, 34], and of course, in the compressive matched filter problem, the unknown signal delay is manifested in the time domain. In hardware, the requisite spectral samples for the compressive matched filter could be acquired by correlating the incoming signal with a bank of oscillators tuned to random frequencies, or by following a Fourier transforming device (such as a SAW processor [18]) with a random sampler. Although the analysis in our paper is limited to one-dimensional signals, one could also envision formulating the matched filtering problem for two-dimensional images, and random samples of a two-dimensional spectrum could be acquired by combining the Fourier transforming property of a lens with a random sampler.

1.2.2 Analytical Framework and Summary of Main Results

In this paper, we develop an analytical framework for studying the compressive matched filter based on tools from probability theory and empirical processes. To help build intuition, in Sections 2.1 through 2.4 we first study the problem fully in the case of noiseless measurements. In Section 2.5 we then extend all of our analysis in parallel fashion to account for measurement noise. Section 4 and several appendices provide supporting proofs for all of our main results.

For both the noiseless and noisy problem formulations, we begin by showing that the output of the correlation-based estimator can be modeled as a random process whose mean equals the scaled, shifted autocorrelation function of the template signal. Noting that the scaled, shifted autocorrelation function of the template signal (the mean of this random process) peaks at τ_0 , we estimate τ_0 by finding the empirical maximum of the random process, and we give guarantees about the accuracy of this estimate by showing that the random process does not vary too much from its mean. Given the estimate of the delay, an estimate of the amplitude A follows easily via least-squares, just as with the standard matched filer.

We approach the analysis as follows. In Theorems 1 and 4, we adapt the proof of the Dudley inequality to show that the expected maximum deviation of this random process from its mean decreases sharply as the number of measurements increases. A bit more formally, Theorem 1 states that in the noiseless case, the expected maximum deviation of this process from its mean decreases roughly like $m^{-1/2}$ (normalized by the peak value of the mean function). Theorem 4 quantifies the amount of additional deviation one would expect based on noise in the observations. In Theorems 2 and 5, we then derive a probabilistic tail bound on the maximum deviation of this process from its mean. Specifically, Theorem 2 guarantees that with high probability the noiseless process stays uniformly close to its mean, and Theorem 5 guarantees that with high probability the maximum additional deviation caused by noise is also bounded. Finally, in Corollaries 3 and 6, we pull these results together to establish bounds on the number of measurements required to guarantee that the empirical maximum of this random process occurs sufficiently close to the true peak of its mean function. Specifically, Corollary 3 ensures in the noiseless case that when the template signal has an autocorrelation function with a single prominent peak, no values of τ far from τ_0 can yield an estimate close to the true peak. Corollary 6 extends this result to account for noise and leads to the central result: the compressive matched filter will successfully estimate τ_0 from m random frequency-domain samples (with high probability, and within a prescribed tolerance) as long as m scales inversely with the signal-to-noise ratio and logarithmically in the observation bandwidth $|\Omega|$ and the possible range of delays $|T|$.

All of our bounds depend on the degree to which the template signal s_0 is concentrated in the frequency domain. As might be expected given the uniform random sampling strategy on Ω , signals whose spectrum is relatively flat across Ω will require the fewest measurements, while signals with highly peaked spectra will require the most. These issues are carefully quantified and discussed in detail throughout Section 2.

1.2.3 Exchanging Time and Frequency

It is important to point out that the roles of time and frequency are completely interchangeable in our analysis. All of our results from Section 2 can therefore be adapted to the “dual” problem of estimating the unknown carrier frequency of a modulated signal given a small number of time-domain samples of that signal; the time domain becomes our observation domain, and the frequency domain becomes the domain in which we wish to determine the unknown shift of the known template signal. For the sake of space, we do not restate all of our results in this context, although the conclusion is clear: a compressive matched filter can successfully estimate an unknown modulation frequency ω_0 from m random time-domain samples (with high probability, and within a prescribed tolerance) as long as m scales inversely with the signal-to-noise ratio and logarithmically in the observation duration $|T|$ and the possible range of carrier frequencies $|\Omega|$. (A Nyquist-based approach, in contrast, would require a sampling rate linearly proportional to $|\Omega|$ but could tolerate somewhat lower signal-to-noise ratios.) The bounds will also depend — in this case — on the degree to which the template signal is concentrated in the time domain. Signals whose envelope is relatively flat across T will require the fewest measurements, while signals with highly peaked envelopes will require the most.

While we do not discuss this problem further in full generality, we do briefly examine a special case, namely the problem of estimating the frequency of a pure sinusoidal tone from noisy time-domain samples. Such a problem is an ideal candidate for the compressive matched filter because a pure sinusoidal tone has a perfectly flat envelope in the time domain. We discuss this tone estimation problem in Section 3, and carefully quantify (in Corollaries 7 and 8) the number of random time-domain samples required to successfully estimate the tone’s frequency. We also address an important practical question: at how many points is it necessary to sample (or query) the random process when searching for its peak? Using an adaptation of Corollary 3, we show for the noiseless case that the empirical peak from a finite set of samples of the random process (with sufficiently dense sampling) must occur within a certain distance of the true peak of the continuous random process. One can therefore employ a grid search strategy for implementing the compressive matched filter, and from the empirical maximum on this grid, one can actually employ a local concave ascent to find the exact value for ω_0 . We close Section 3 with a stylized application illustrating the potential of extending this work to the problem of determining the arrival time of a linear chirp.

1.3 Related Work

To the best of our knowledge, our framework for studying the compressive matched filter is novel. Prior statistical analysis for compressive inference problems has focused specifically on problems of signal detection or classification from a finite model set [11, 16, 17] or employed a geometric point-of-view based on a stable embedding of signal family from an original finite-dimensional signal space into a lower-dimensional measurement space [6, 7]. Our work takes a substantially different approach, considering the inference of a continuous-valued shift parameter from a continuous-time received signal, and more thoroughly characterizing the statistics of the problem using the language and tools of empirical processes.

As mentioned above, similar probabilistic tools have been employed in CS, but for the analysis of the sparse signal recovery problem [28, 30, 32–34]. While in principle one could

view the matched filter problem as that of recovering a 1-sparse signal from a dictionary $\{s_0(t - \tau) : \tau \in T\}$ of possible candidates, such a dictionary would have infinite size and extremely high coherence, preventing the application of most standard recovery analysis techniques. One recent work [14] has formalized the matched filter problem using signal recovery principles and a union of subspaces model. However, this work is quite different from ours in that it does not theoretically study noise sensitivity and relies on a non-random sampling architecture that is carefully designed to facilitate the solution of the recovery problem. Interestingly, outside the field of CS, very similar random processes to those that we study have also arisen in the analysis of the spectral norm of random Toeplitz matrices [27].

The second part of this paper adapts our analysis of the compressive matched filter to the problem of estimating the frequency of a pure sinusoidal tone from a small number of random time-domain samples. The recovery of signals that are sparse in the frequency domain based on compressive measurements is a problem that has been well-studied in the CS literature, although most work in this area has been concerned with signals that can be written as trigonometric polynomials [5, 15, 20, 35]. Some techniques for recovering off-grid frequency-sparse signals have been proposed that involve windowing [35] or other classical techniques from the field of spectral estimation [10], and other work has considered the more general problem of recovering continuous-time signals based on a union of subspaces model [13], but the analysis that we present is more sharply focused on the statistics of the simpler pure tone estimation problem.

Finally, we would like to point out some of the differences between the tone estimation problem considered in this paper and the classical problem of estimating the power spectrum of a random process from samples at random locations (see [2, 3, 24, 26]). In Sections 2 and 3, we will show how the output of the compressive matched filter is a random process whose mean is the template autocorrelation function. This random process is completely specified by the samples we have observed, and rather than merely estimating its second-order statistics, we will be interested in establishing a uniform bound on its deviation from the template; this will allow us to conclude that it peaks at or near the correct location. It is also worth mentioning that our work differs from Rife and Boorstyn’s classical analysis of the single-tone parameter estimation problem [31]. Specifically, our work permits sampling below the Nyquist rate, and with high probability we provide an absolute bound on the accuracy of the frequency estimate, rather than involving the Cramér-Rao bound.

2 Analytical Framework and Main Results

2.1 Problem Statement

2.1.1 Signal Model

Suppose we have received a signal $A \cdot s_0(t - \tau_0)$, where $s_0(t)$ is a known signal template, and τ_0 and A are the unknown delay and amplitude, respectively. We assume that the unknown delay τ_0 (also called the *time-of-arrival*) is restricted to some interval $T = [\tau_{\min}, \tau_{\max}] \subset \mathbb{R}$. We make no particular assumptions about s_0 , although our bounds will depend on the properties of s_0 over the range of frequencies where it is observed.

We will consider two closely related cases in this paper: in the *real case*, we restrict both s_0 and A to be real-valued, whereas in the *complex case*, we allow both s_0 and A to be complex-valued. Much of our analysis will be identical for the real and complex cases, but we will specialize our discussions to distinguish between the two cases when necessary.

2.1.2 Observations

We would like to estimate τ_0 and A based on random samples of the Fourier transform of the received signal. In particular, we suppose that we acquire m samples of the Fourier transform of $A \cdot s_0(t - \tau_0)$ at frequencies $\omega_1, \omega_2, \dots, \omega_m$, which are drawn independently at random from a uniform distribution on some interval $\Omega = [-\omega_{\max}, \omega_{\max}]$ in the frequency domain. Typically, one would choose Ω roughly equal to the essential bandwidth of s_0 , although this is not strictly necessary; we more carefully discuss the implications of choosing Ω in Section 2.5 below.

The vector of observations $y \in \mathbb{C}^m$ is formed as

$$y[k] = A \int_{-\infty}^{\infty} s_0(t - \tau_0) e^{-i\omega_k t} dt = A \cdot e^{-i\omega_k \tau_0} \hat{s}_0(\omega_k), \quad k = 1, 2, \dots, m, \quad (1)$$

where $\hat{s}_0(\omega)$ denotes the Fourier transform of $s_0(t)$. For the moment, we assume that these observations are noiseless. In Section 2.5 we extend our formulation to account for measurement noise.

We define $s(t)$ to be a low-pass filtered version of $s_0(t)$ having frequency content bandlimited to the interval Ω . More formally,

$$s(t) := \frac{1}{2\pi} \int_{\Omega} \hat{s}_0(\omega) e^{i\omega t} d\omega.$$

It follows that $\hat{s}(\omega) = \hat{s}_0(\omega)$ for all $\omega \in \Omega$ and that $\hat{s}(\omega) = 0$ for all $\omega \notin \Omega$. Thus, because our Fourier-domain observations are limited to the interval Ω , we may rewrite the expression (1) for our observations as

$$y[k] = A \cdot e^{-i\omega_k \tau_0} \hat{s}(\omega_k), \quad k = 1, 2, \dots, m.$$

Consequently, all of our subsequent analysis will depend only on properties of the bandlimited signal $s(t)$.

2.1.3 Least-Squares Estimation

Given the observation vector y , a natural approach to estimating τ_0 and A is to find the delay and amplitude which best explain the measurements in a least-squares sense. (Such a least-squares estimate coincides with the maximum likelihood estimate in the case of Gaussian measurement noise, as we consider in Section 2.5.) More formally, we define

$$(\hat{\tau}_0, \hat{A}) := \arg \min_{\tau, A} \sum_{k=1}^m |y[k] - A \cdot e^{-i\omega_k \tau} \hat{s}(\omega_k)|^2 = \arg \min_{\tau, A} \|y - A\psi_{\tau}\|_2^2, \quad (2)$$

where for any $\tau \in T$, the test vector $\psi_\tau \in \mathbb{C}^m$ is given by:

$$\psi_\tau[k] = e^{-i\omega_k \tau} \hat{s}(\omega_k), \quad k = 1, 2, \dots, m.$$

For a given estimate $\hat{\tau}_0$ of the delay, one can derive a closed form expression for the amplitude A that minimizes (2):

$$\hat{A} = \begin{cases} \frac{\operatorname{Re}\langle y, \psi_{\hat{\tau}_0} \rangle}{\|\psi_{\hat{\tau}_0}\|_2^2}, & \text{real case,} \\ \frac{\langle y, \psi_{\hat{\tau}_0} \rangle}{\|\psi_{\hat{\tau}_0}\|_2^2}, & \text{complex case.} \end{cases} \quad (3)$$

Plugging (3) in to (2), we see that the optimal time-of-arrival estimate is given by

$$\hat{\tau}_0 = \begin{cases} \arg \min_{\tau \in T} \frac{-|\operatorname{Re}\langle y, \psi_\tau \rangle|^2}{\|\psi_\tau\|_2^2}, & \text{real case,} \\ \arg \min_{\tau \in T} \frac{-|\langle y, \psi_\tau \rangle|^2}{\|\psi_\tau\|_2^2}, & \text{complex case.} \end{cases}$$

Finally, noting that $\|\psi_\tau\|$ is constant over all $\tau \in T$, we obtain a simplified expression for the least-squares estimate of τ_0 :

$$\hat{\tau}_0 = \begin{cases} \arg \max_{\tau \in T} |\operatorname{Re}\langle y, \psi_\tau \rangle|, & \text{real case,} \\ \arg \max_{\tau \in T} |\langle y, \psi_\tau \rangle|, & \text{complex case.} \end{cases} \quad (4)$$

Equation (4) suggests a correlation-based strategy for estimating τ_0 ; this strategy is a natural generalization of the traditional time-domain ‘‘matched filter’’ to our measurement setting. For this reason, we refer to such an estimator as a *compressive matched filter*, and our focus in this paper will be on the accuracy with which τ_0 can be estimated using such an estimator. Because \hat{A} is subsequently defined in terms of $\hat{\tau}_0$, one could easily extend our analysis to bound the accuracy of estimating A .

2.2 Noiseless Analysis

In order to study the performance of a correlation-based estimator for τ_0 , let us define the complex-valued random process $X(\tau)$ on T to be the correlation of the observations y with each of the test vectors ψ_τ ,

$$X(\tau) := \langle y, \psi_\tau \rangle = A \sum_{k=1}^m |\hat{s}(\omega_k)|^2 e^{i\omega_k(\tau-\tau_0)}. \quad (5)$$

This random process has mean function

$$\mathbb{E}X(\tau) = A \sum_{k=1}^m \mathbb{E} |\hat{s}(\omega_k)|^2 e^{i\omega_k(\tau-\tau_0)} = \frac{Am}{|\Omega|} \int_{\Omega} |\hat{s}(\omega)|^2 e^{i\omega(\tau-\tau_0)} d\omega = \frac{2\pi Am}{|\Omega|} R_{ss}(\tau - \tau_0), \quad (6)$$

where $R_{ss}(\cdot) = (s(t) \star s^*(-t))(\cdot)$ denotes the autocorrelation function of $s(t)$.

In the complex case, the compressive matched filter estimate (4) for τ_0 can be interpreted as a search for the maximizer of $|X(\tau)|$. Because $|R_{ss}(\cdot)|$ is maximized at the origin, one would expect informally that, on average, finding the maximum magnitude of the process $X(\tau)$ should correctly estimate τ_0 . In the real case, the compressive matched filter estimate for τ_0 can be interpreted as a search for the maximizer of $|\operatorname{Re}(X(\tau))|$. However, in this case

we note that since $R_{ss}(\cdot)$ is real, we will have $\mathbb{E} \operatorname{Re}(X(\tau)) = \operatorname{Re}(\mathbb{E} X(\tau)) = \mathbb{E} X(\tau)$, which again has magnitude maximized at τ_0 , and so informally, finding the maximum magnitude of the process $\operatorname{Re}(X(\tau))$ should correctly estimate τ_0 .

An equivalent, and potentially more revealing, way to frame the delay estimation problem is to observe that we could rescale $X(\tau)$ to obtain an estimate of the ideal autocorrelation function $R_{ss}(\cdot)$ (up to the unknown amplitude A and translation τ_0). Define

$$\tilde{R}_{ss}(\tau) := \begin{cases} \frac{|\Omega|}{2\pi m} \cdot \operatorname{Re}(X(\tau)), & \text{real case,} \\ \frac{|\Omega|}{2\pi m} \cdot X(\tau), & \text{complex case.} \end{cases} \quad (7)$$

One way to interpret this estimate is that we have approximated the scaled, shifted autocorrelation function, $AR_{ss}(\tau - \tau_0)$, as a discrete sum with samples taken at random locations in Ω ; equation (6) tell us that this estimate is unbiased, since $\mathbb{E}[\tilde{R}_{ss}(\tau)] = AR_{ss}(\tau - \tau_0)$.

It is clear that solving the least-squares problem (4) is equivalent to finding the maximum of $|\tilde{R}_{ss}(\tau)|$. Our main concern will be quantifying how close the random process $\tilde{R}_{ss}(\tau)$ is to its mean. It is worth noting that, if the measurements are perfectly clean and we are able to perform all computations to infinite precision, then $|\tilde{R}_{ss}(\tau)|$ is actually guaranteed to peak at τ_0 , where it takes its maximum value of $|\Omega|(|A|2\pi m)^{-1}\|y\|_2^2$. But what Theorems 1 and 2 will tell us is that if m is large enough, then there will be a tangible gap between this peak at τ_0 and the values of $|\tilde{R}_{ss}(\tau)|$ for all τ bounded some distance away from τ_0 . As we will then see in Section 2.5, this gap will make the maximizer of $|\tilde{R}_{ss}|$ a robust estimate of τ_0 in the presence of noise.

To simplify some of the notation, we will use η to denote the peak magnitude of the mean function $AR_{ss}(\tau - \tau_0)$,

$$\eta = |AR_{ss}(0)| = |A| \|s\|_2^2 = |A| \frac{\|\hat{s}\|_2^2}{2\pi}.$$

Our results will depend on how concentrated the Fourier transform $\hat{s}(\omega)$ is over the sampling domain Ω . Intuitively, if \hat{s} is spread out more or less evenly over Ω , then each sample will give us some information about the return signal. If \hat{s} is concentrated on a small set within Ω , then only a small number of the randomly chosen samples will tell us anything at all. We will quantify this concentration in two different ways. We introduce

$$\mu_1 = \frac{\sqrt{|\Omega|} \cdot \|\hat{s}\|_4^2}{\|\hat{s}\|_2^2}, \quad \text{and} \quad \mu_2 = \frac{|\Omega| \cdot \|\hat{s}\|_\infty^2}{\|\hat{s}\|_2^2}.$$

If the energy of \hat{s} is equally spread over the sampling domain Ω , that is if $|\hat{s}(\omega)| = |\Omega|^{-1/2}\|\hat{s}\|_2$ for all $\omega \in \Omega$, then it is easy to see that $\mu_1 = \mu_2 = 1$. If most of the energy in $\hat{s}(\omega)$ is concentrated on a small subset of Ω , then μ_1 and μ_2 will be large (and in fact, they can be made arbitrarily large).

We start by getting a rough idea of how close $\tilde{R}_{ss}(\tau)$ is to its mean by looking at the variance at a shift τ . Since $|\operatorname{Re}(X(\tau))| \leq |X(\tau)|$, we can bound the second moment of $\tilde{R}_{ss}(\tau)$ in both

the real and complex cases:

$$\begin{aligned}
\mathbb{E} \left[|\tilde{R}_{ss}(\tau)|^2 \right] &\leq \frac{|\Omega|^2}{4\pi^2 m^2} \mathbb{E} |X(\tau)|^2 \\
&= \frac{|A|^2 |\Omega|^2}{4\pi^2 m^2} \sum_{k_1=1}^m \sum_{k_2=1}^m \mathbb{E} \left[|\hat{s}(\omega_{k_1})|^2 |\hat{s}(\omega_{k_2})|^2 e^{-i\omega_{k_1}(\tau-\tau_0)} e^{i\omega_{k_2}(\tau-\tau_0)} \right] \\
&= \frac{|A|^2 |\Omega|^2}{4\pi^2 m^2} \sum_{k_1=1}^m \mathbb{E} [|\hat{s}(\omega_{k_1})|^4] \\
&\quad + \frac{|A|^2 |\Omega|^2}{4\pi^2 m^2} \sum_{k_1=1}^m \sum_{k_2 \neq k_1} \mathbb{E}_{\omega_{k_1}} \left[|\hat{s}(\omega_{k_1})|^2 e^{-i\omega_{k_1}(\tau-\tau_0)} \right] \mathbb{E}_{\omega_{k_2}} \left[|\hat{s}(\omega_{k_2})|^2 e^{i\omega_{k_2}(\tau-\tau_0)} \right] \\
&= \frac{|A|^2 |\Omega|^2}{4\pi^2 m^2} \sum_{k_1=1}^m \frac{1}{|\Omega|} \int_{\Omega} |\hat{s}(\omega)|^4 d\omega \\
&\quad + \frac{|A|^2 |\Omega|^2}{4\pi^2 m^2} \sum_{k_1=1}^m \sum_{k_2 \neq k_1} \frac{1}{|\Omega|^2} \left(\int_{\Omega} |\hat{s}(\omega)|^2 e^{i\omega(\tau-\tau_0)} d\omega \right)^* \left(\int_{\Omega} |\hat{s}(\omega)|^2 e^{i\omega(\tau-\tau_0)} d\omega \right) \\
&= \frac{|A|^2 |\Omega|^2}{4\pi^2 m^2} \left(\frac{m}{|\Omega|} \|\hat{s}\|_4^4 + \frac{m(m-1)4\pi^2}{|\Omega|^2} |R_{ss}(\tau - \tau_0)|^2 \right) \\
&= \frac{|A|^2 |\Omega| \|\hat{s}\|_4^4}{4\pi^2 m} + \frac{m-1}{m} |A|^2 |R_{ss}(\tau - \tau_0)|^2,
\end{aligned}$$

whereas $|\mathbb{E}[\tilde{R}_{ss}(\tau)]|^2 = |A|^2 |R_{ss}(\tau - \tau_0)|^2$. Therefore,

$$\text{Var} \left[\tilde{R}_{ss}(\tau) \right] = \mathbb{E} \left[|\tilde{R}_{ss}(\tau)|^2 \right] - \left| \mathbb{E}[\tilde{R}_{ss}(\tau)] \right|^2 \leq \frac{\eta^2}{m} \left(\mu_1^2 - \frac{4\pi^2 |R_{ss}(\tau - \tau_0)|^2}{\|\hat{s}\|_2^4} \right) \leq \frac{\eta^2 \mu_1^2}{m}.$$

Using Jensen's inequality, we then obtain a bound for the expected deviation of $\tilde{R}_{ss}(\tau)$ from its mean at a fixed shift τ :

$$\mathbb{E} \left| \tilde{R}_{ss}(\tau) - A R_{ss}(\tau - \tau_0) \right| \leq \frac{\eta \mu_1}{\sqrt{m}}. \quad (8)$$

As expected, this deviation gets smaller as m increases, and scales with μ_1 .

Our first theorem gives us a *uniform* bound for the expected maximum deviation of $\tilde{R}_{ss}(\tau)$ from its mean over all $\tau \in T$. The following result is proved in Section 4.1.

Theorem 1. Suppose that $|\Omega||T| \geq 3$.¹ Then the autocorrelation function estimate $\tilde{R}_{ss}(\tau)$, as defined in (7), obeys

$$\begin{aligned}
\mathbb{E} \sup_{\tau \in T} \left| \tilde{R}_{ss}(\tau) - A R_{ss}(\tau - \tau_0) \right| &\leq \frac{\eta \mu_1}{\sqrt{m}} \cdot \left(4.25 \sqrt{\log(2|\Omega||T|)} + 2.28 \right) \\
&\leq 5.96 \cdot \frac{\eta \mu_1}{\sqrt{m}} \cdot \sqrt{\log(2|\Omega||T|)}. \quad (9)
\end{aligned}$$

The essential difference between (8) and (9) is the factor of $\sqrt{\log(2|\Omega||T|)}$ — this is the price we are paying for a bound which holds uniformly over all $\tau \in T$. The bound slowly

¹If $|\Omega||T| < 3$, this theorem and all of our bounds still hold but with weaker constants.

loosens as the time-bandwidth product $|T||\Omega|$ gets larger; this effect is weak but necessary, as $|T||\Omega|$ affects the complexity of the random process. (A similar penalty arises in standard CS bounds [5, 9], where the number of measurements required for successful, robust signal recovery grows logarithmically with the ambient dimension of the signal space — this logarithmic dependence is known to be sharp.)

We note that Theorem 1 could be proved using the Dudley inequality [12], a classical tool which relates the supremum of a random process to the geometry of its index set; the main challenges arise in computing covering numbers for the index set T under certain metrics defined in terms of the random process $\tilde{R}_{ss}(\tau)$ and in adapting the Dudley argument to complex numbers. To provide better insight and to obtain sharper constants, however, our proof in Section 4.1 more directly customizes the derivation of the Dudley inequality for our particular scenario. We also note that a simple application of the Sudakov minoration principle [23] (after computing the necessary metrics) reveals that the bound in Theorem 1 is indeed sharp (up to a constant). Intuitively, $|\Omega||T|$ is the number of points on a grid of resolution $1/|\Omega|$ on T necessary to control the deviations of the random process.

Theorem 1 demonstrates that $\tilde{R}_{ss}(\tau)$ is close to its mean in expectation; our second theorem demonstrates that it is also close with high probability. The following is proved in Section 4.1.

Theorem 2. *Fix $\delta > 0$ and let*

$$U = C_1 \cdot \max \left(\frac{\eta\mu_1}{\sqrt{m}}, \frac{\eta\mu_2}{m} \cdot \sqrt{\log(4/\delta)} \right) \cdot \sqrt{\log(12|\Omega||T|/\delta)}, \quad (10)$$

where C_1 is a known universal constant. If $|\Omega||T| \geq 3$, then the autocorrelation function estimate $\tilde{R}_{ss}(\tau)$, as defined in (7), obeys

$$\mathbb{P} \left\{ \sup_{\tau \in T} \left| \tilde{R}_{ss}(\tau) - AR_{ss}(\tau - \tau_0) \right| > U \right\} \leq \delta. \quad (11)$$

2.3 Example: A Gaussian Pulse

A concrete example will help illustrate what Theorems 1 and 2 are telling us about the effectiveness of the compressive matched filter. Suppose that $s_0(t)$ is a real-valued Gaussian pulse with unit energy,

$$s_0(t) = \pi^{-1/4} a^{-1/2} e^{-t^2/2a^2}, \quad (12)$$

We will assume that this pulse is received with a time-of-arrival τ_0 in the interval $T = [0, 1]$, and that it is scaled by an unknown real-valued amplitude A . We will also assume that the width a of the pulse is much less than 1, and so to estimate τ_0 reliably from samples in the time domain, we would need on the order of $1/a$ samples on T . Figure 1(a) shows an example received signal $A \cdot s_0(t - \tau_0)$ for $A = 1$, $\tau_0 = 0.4$ and $a = 1/200$.

The Fourier transform of s_0 is

$$\hat{s}_0(\omega) = \sqrt{2a}\pi^{1/4} e^{-a^2\omega^2/2}.$$

We will take as our sampling domain $\Omega = [-3/a, 3/a]$; \hat{s} is simply \hat{s}_0 bandlimited to this interval. The bandlimited signal \hat{s} is nearly identical to \hat{s}_0 ; we can calculate

$$\|\hat{s}_0\|_2^2 = 2\pi, \quad \|\hat{s}_0\|_4^2 = (2\pi)^{3/4} \sqrt{a},$$

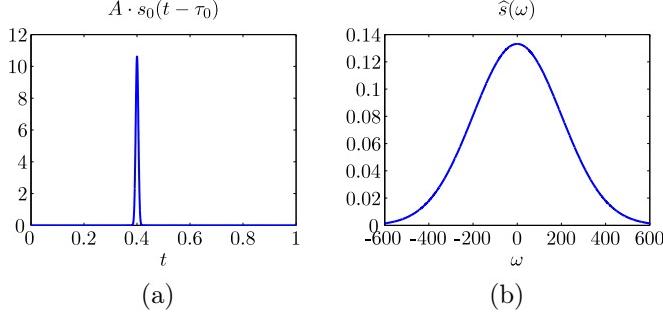


Figure 1: (a) The return signal for Gaussian pulse from (12) with $a = 1/200$ and return parameters $A = 1$ and $\tau_0 = 0.4$. (b) The Fourier transform $\hat{s}(\omega)$ on $\Omega = [-600, 600]$. Since \hat{s} is relatively diffuse over Ω , both measures of frequency concentration are not too large: $\mu_1 \leq 1.6$, $\mu_2 \leq 3.4$.

and standard bounds for integrating the tails of $e^{-a^2\omega^2}$ show us that $\|\hat{s}\|_2^2$ is within $1.4 \cdot 10^{-4}$ of $\|\hat{s}_0\|_2^2$ and $\|\hat{s}\|_4^2$ is within 10^{-8} of $\|\hat{s}_0\|_4^2$. We can safely say that $\mu_1 \leq 1.6$ and $\mu_2 \leq 3.4$ (these values are the same for all choices of a). The Fourier transform for $a = 1/200$ over the range $\Omega = [-600, 600]$ is shown in Figure 1(b).

For a received signal with parameters $A = 1$, $\tau_0 = 0.4$, and $a = 1/200$, Figure 2 shows the estimate $\tilde{R}_{ss}(\tau)$ of the scaled, shifted autocorrelation function based on $m = 10, 20$, and 50 random frequency domain samples, along with the true scaled, shifted autocorrelation function $A \cdot R_{ss}(\tau - \tau_0) \approx e^{-(\tau - \tau_0)^2/(4a^2)}$. In all cases, we see that $|\tilde{R}_{ss}(\tau)|$ reaches its peak exactly at τ_0 ; as noted in Section 2.2, this is to be expected in the case of noiseless measurements. However, we also see a gap between the peak at τ_0 and the remainder of the estimate that becomes larger as m increases.

To see how this behavior is supported by our theory, note that for a Gaussian pulse with $A = 1$, we know that the mean function $R_{ss}(\tau - \tau_0) = 1$ for $\tau = \tau_0$ and $R_{ss}(\tau - \tau_0) \leq 0.1054$ for $|\tau - \tau_0| > 3a$. (For simplicity, these calculations assume $s = s_0$ exactly.) If U is the value from (10) in Theorem 2, then we are guaranteed that the difference between the peak value $\tilde{R}_{ss}(\tau_0)$ and any $\tilde{R}_{ss}(\tau)$ for $|\tau - \tau_0| \geq 3a$ is at least ϵ when $1 - U \geq 0.1054 + U + \epsilon$, i.e., when $U \leq \frac{0.8946 - \epsilon}{2}$. Note that U can be made small enough for large enough m , namely $m \gtrsim \log(1/a)$. Thus using the compressive matched filter, we can reliably infer the time-of-arrival from $\sim \log(1/a)$ randomly chosen samples in the frequency domain as opposed to $\sim 1/a$ equally spaced samples in the time domain.

2.4 General Noiseless Performance Characterization

Our statements about quantifying the number of samples needed to ensure a clear separation between the peak of $|\tilde{R}_{ss}(\tau)|$ and the function away from the peak are easily generalized. The statements in this section can be interpreted as a condition on the number of samples needed to ensure the successful operation of the compressive matched filter. The result below is interesting when the underlying autocorrelation function $R_{ss}(\tau)$ has one main peak (a “main lobe”) centered at $\tau = 0$, and is relatively small away from the origin. This situation is typical, but similar statements could be formulated depending on the assumptions one wishes to impose on $s(t)$ (and its autocorrelation function).

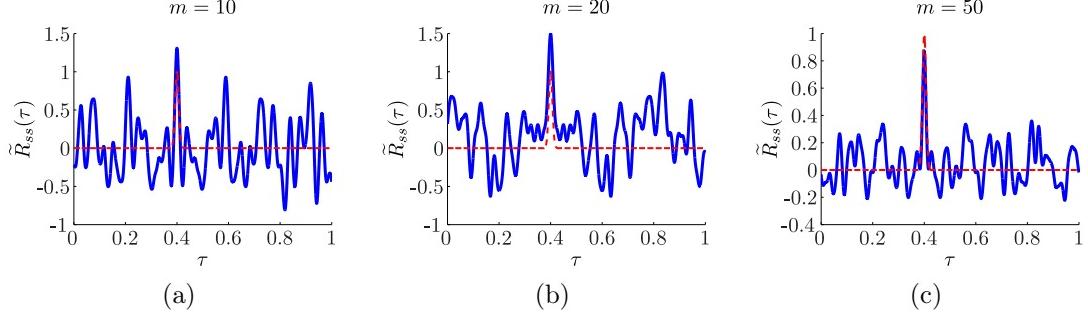


Figure 2: Estimated scaled, shifted autocorrelation function $\tilde{R}_{ss}(\tau)$ (solid blue line) and true scaled, shifted autocorrelation function $A \cdot R_{ss}(\tau - \tau_0)$ (dashed red line) for (a) $m = 10$, (b) $m = 20$, (c) and $m = 50$ measurements. $\tilde{R}_{ss}(\tau)$ is a random process whose mean is $A \cdot R_{ss}(\tau - \tau_0)$; as the number of measurements increases, this process deviates less from its mean (see Theorems 1 and 2).

Corollary 3. Suppose there exist constants $\alpha_1 \in [0, 1)$ and $\alpha_2 > 0$ such that $|R_{ss}(\tau)| \leq \alpha_1 R_{ss}(0)$ for all $|\tau| > \alpha_2$, and choose $\epsilon \in [0, 1 - \alpha_1]$. Suppose also that $|\Omega| |T| \geq 3$ and that

$$m > C_2 \cdot \max \left(\frac{\log(12|\Omega||T|/\delta)}{(1 - \alpha_1 - \epsilon)^2} \cdot \mu_1^2, \frac{\sqrt{\log(4/\delta) \log(12|\Omega||T|/\delta)}}{1 - \alpha_1 - \epsilon} \cdot \mu_2 \right), \quad (13)$$

where C_2 is a known universal constant. Then with probability at least $1 - \delta$, $|\tilde{R}_{ss}(\tau_0)| > |\tilde{R}_{ss}(\tau)| + \epsilon\eta$ for all τ such that $|\tau - \tau_0| > \alpha_2$.

Proof. Supposing we have the concentration suggested by (11), we will have $|\tilde{R}_{ss}(\tau_0)| \geq |A|R_{ss}(0) - U$ and $|\tilde{R}_{ss}(\tau)| \leq \alpha_1|A|R_{ss}(0) + U$ for all τ such that $|\tau - \tau_0| > \alpha_2$. If (13) is satisfied with $C_2 = \max(4C_1^2, 2C_1)$, then $U < \frac{1}{2}|A|R_{ss}(0)(1 - \alpha_1 - \epsilon)$ and it follows that $|A|R_{ss}(0) - U > \alpha_1|A|R_{ss}(0) + U + \epsilon|A|R_{ss}(0)$. A slightly stronger version of this corollary also holds if one omits C_2 and chooses constants of $4C_1^2$ and $2C_1$ for the first and second terms in (13), respectively. \square

For the case of noiseless measurements, Corollary 3 ensures that no values of τ far from τ_0 can give $|\tilde{R}_{ss}(\tau)|$ close to $|\tilde{R}_{ss}(\tau_0)|$. This behavior will become particularly relevant in Section 2.5, where we introduce noise into the measurement process.

We can reveal some of the intuition behind the measurement bound (13) by considering three special cases for the signal s . First, consider $s(t)$ for which $|\hat{s}(\omega)|$ is uniform over Ω . In this case, we have $|\hat{s}(\omega)| = |\Omega|^{-1/2}\|\hat{s}\|_2$ for all $\omega \in \Omega$ and so $\mu_1 = \mu_2 = 1$. This means that the requisite number of random measurements (13) for successful operation of the compressive matched filter scales as $m \sim \log(|\Omega||T|)$.

Alternatively, consider the case where $|\hat{s}(\omega)|$ is not perfectly uniform over Ω , but rather we assume that for some $\beta \geq 1$ it obeys $|\hat{s}(\omega)| \leq \beta|\Omega|^{-1/2}\|\hat{s}\|_2$ for all $\omega \in \Omega$, and so $\mu_2 \leq \beta^2$. Using the fact that $\|\hat{s}\|_4^4 \leq \|\hat{s}\|_2^2\|\hat{s}\|_\infty^2$ gives us the estimate $\mu_1 \leq \beta$. Therefore, (13) now demands that $m \sim \beta^2 \log(|\Omega||T|)$ — the factor of β^2 is the price we pay for the non-uniformity of \hat{s} .

As a final example, consider the special case where s is bandlimited to some interval $\Omega_B \subseteq \Omega$

and $|\widehat{s}(\omega)|$ is uniform over Ω_B , i.e.,

$$|\widehat{s}(\omega)| = \begin{cases} \frac{\|\widehat{s}\|_2}{\sqrt{|\Omega_B|}}, & \omega \in \Omega_B \\ 0, & \omega \in \Omega \setminus \Omega_B, \end{cases}$$

One way we could interpret this situation is that we have chosen the sampling domain Ω to be too large. In this case, we have $\mu_1 = \sqrt{|\Omega|/|\Omega_B|}$ and $\mu_2 = |\Omega|/|\Omega_B|$, and so (13) now demands that $m \sim (|\Omega|/|\Omega_B|) \log(|\Omega||T|)$. The penalty $|\Omega|/|\Omega_B|$ is a natural oversampling factor since, on average, only $|\Omega_B|$ out of every $|\Omega|$ random Fourier samples will carry any information about the signal.

2.5 Robustness in the Presence of Measurement Noise

We now extend our analysis to account for additive complex-valued noise in our observations. For random frequencies $\{\omega_k\}$ taken uniformly from Ω , we assume that the noisy measurement vector, y_n , is formed as

$$y_n[k] = Ae^{-i\omega_k \tau_0} \widehat{s}_0(\omega_k) + n_k, \quad k = 1, 2, \dots, m,$$

where the additive noise terms $\{n_k\}$ are independent zero-mean complex-valued Gaussian random variables² with variance σ_n^2 , and the noise vector is $n := [n_1, n_2, \dots, n_m]^T$. Computing the inner product of y_n with the test vector ψ_τ for all $\tau \in T$ leads us to the process

$$X_n(\tau) := \langle y_n, \psi_\tau \rangle = X(\tau) + N(\tau), \quad (14)$$

where $X(\tau)$ is as defined in Section 2.2, and $N(\tau)$ is the noise process that quantifies the effect of additive noise in our analysis:

$$N(\tau) := \langle n, \psi_\tau \rangle = \sum_{k=1}^m n_k \widehat{s}_0^*(\omega_k) e^{i\omega_k \tau} = \sum_{k=1}^m n_k \widehat{s}^*(\omega_k) e^{i\omega_k \tau}.$$

The noise process is zero-mean, i.e., $\mathbb{E}N(\tau) = 0$.

Thus, in the case of noisy observations, we can estimate the ideal autocorrelation function $R_{ss}(\cdot)$ (up to the unknown amplitude A and translation τ_0) simply by rescaling the noisy random process $X_n(\tau)$. Let us define

$$\widetilde{N}(\tau) := \begin{cases} \frac{|\Omega|}{2\pi m} \cdot \text{Re}(N(\tau)), & \text{real case,} \\ \frac{|\Omega|}{2\pi m} \cdot N(\tau), & \text{complex case,} \end{cases} \quad (15)$$

and note that in either case, $\mathbb{E}[\widetilde{N}(\tau)] = 0$. Then, if we set

$$\widetilde{R}_{ss,n}(\tau) := \begin{cases} \frac{|\Omega|}{2\pi m} \cdot \text{Re}(X_n(\tau)), & \text{real case,} \\ \frac{|\Omega|}{2\pi m} \cdot X_n(\tau), & \text{complex case,} \end{cases}$$

it follows in either case that

$$\widetilde{R}_{ss,n}(\tau) = \widetilde{R}_{ss}(\tau) + \widetilde{N}(\tau), \quad (16)$$

²That is, the real and imaginary parts of each n_k are independent, real-valued zero-mean Gaussian random variables with variance $\frac{\sigma_n^2}{2}$.

where $\tilde{R}_{ss}(\tau)$ is as defined in (7). This function provides an unbiased estimate of the shifted, scaled autocorrelation function of $s(t)$, since $\mathbb{E}[\tilde{R}_{ss,n}(\tau)] = AR_{ss}(\tau - \tau_0)$.

We can gain some intuition for how the noise is hindering the estimation process with a quick estimate on its expected size at a fixed point τ . In both the real and complex cases, the variance of $\tilde{N}(\tau)$ is bounded by

$$\begin{aligned}\text{Var} [\tilde{N}(\tau)] &\leq \frac{|\Omega|^2}{4\pi^2 m^2} \mathbb{E} \left| \sum_{k=1}^m n_k \hat{s}^*(\omega_k) e^{i\omega_k \tau} \right|^2 \\ &= \frac{|\Omega|^2}{4\pi^2 m^2} \sum_{k_1=1}^m \sum_{k_2=1}^m \mathbb{E}[n_{k_1} n_{k_2}^*] \mathbb{E}[\hat{s}^*(\omega_{k_1}) \hat{s}(\omega_{k_2}) e^{i(\omega_{k_1} - \omega_{k_2})\tau}] \\ &= \frac{|\Omega|^2}{4\pi^2 m^2} \sum_{k=1}^m \mathbb{E}|n_k|^2 \mathbb{E}|\hat{s}(\omega_k)|^2 \\ &= \frac{|\Omega|}{4\pi^2 m} \sigma_n^2 \|\hat{s}\|_2^2,\end{aligned}$$

and so, using Jensen's inequality, we obtain

$$\mathbb{E}|\tilde{N}(\tau)| \leq \frac{1}{2\pi} \cdot \sigma_n \cdot \frac{\sqrt{|\Omega|} \|\hat{s}\|_2}{\sqrt{m}}. \quad (17)$$

Recall that the peak of the noiseless estimate $|\tilde{R}_{ss}(\tau)|$ is on the order of $|A|R_{ss}(0) = |A|(2\pi)^{-1}\|\hat{s}\|_2^2$. Thus the noise process will overwhelm the peak of the noiseless estimate when

$$\sigma_n \sim |A| \|\hat{s}\|_2 \sqrt{\frac{m}{|\Omega|}}. \quad (18)$$

Theorems 4 and 5 below show that for m large enough, we will have essentially the same bound as (17) hold *uniformly* over the entire search interval T . As a result, the amount of noise (size of σ_n^2) the compressive matched filter can withstand is essentially (to within constant and log factors) the same as in (18).

We start with a bound on the expected maximum of the noise process. The following result is proved in Section 4.2.

Theorem 4. *Suppose that $|\Omega||T| \geq 3$. Then the noise process $\tilde{N}(\tau)$, as defined in (15), obeys*

$$\begin{aligned}\mathbb{E} \sup_{\tau \in T} |\tilde{N}(\tau)| &\leq \sigma_n \cdot \frac{\sqrt{|\Omega|} \|\hat{s}\|_2}{\sqrt{m}} \cdot \left(0.199 \sqrt{\log(|\Omega||T|)} + 0.166 \right) \\ &\leq 0.36 \cdot \sigma_n \cdot \frac{\sqrt{|\Omega|} \|\hat{s}\|_2}{\sqrt{m}} \cdot \sqrt{\log(|\Omega||T|)}.\end{aligned}$$

The next theorem shows that, given m large enough, the maximum of the noise process will not be too much larger than its mean with high probability. The following result is also proved in Section 4.2.

Theorem 5. *Fix $\delta > 0$. Suppose that $|\Omega||T| \geq 3$ and that*

$$m \geq C_3 \cdot \max(\mu_1^2, \mu_2) \cdot \log(1/\delta), \quad (19)$$

where C_3 is a known universal constant. Then the noise process $\tilde{N}(\tau)$, as defined in (15), obeys

$$\mathrm{P} \left\{ \sup_{\tau \in T} \left| \tilde{N}(\tau) \right| \geq C_4 \cdot \sigma_n \cdot \frac{\sqrt{|\Omega|} \|\hat{s}\|_2}{\sqrt{m}} \cdot \max \left(\sqrt{\log(|\Omega||T|)}, \sqrt{\log(2/\delta)} \right) \right\} \leq \delta, \quad (20)$$

where C_4 is a known universal constant.

These two theorems, taken in conjunction with Theorems 1 and 2, give us a bound on how far the estimate $\tilde{R}_{ss,n}(\tau)$ created from noisy samples will vary from its mean. With high probability, we will have

$$\left| \tilde{R}_{ss,n}(\tau) - AR_{ss}(\tau - \tau_0) \right| \lesssim \max \left(\frac{\eta\mu_1}{\sqrt{m}}, \frac{\eta\mu_2}{m}, \frac{\sigma_n \sqrt{|\Omega|} \|\hat{s}\|_2}{\sqrt{m}} \right) \cdot \sqrt{\log(|\Omega||T|)}.$$

Just as in the noiseless case, the bound on this deviation of our estimate of the autocorrelation function can be translated directly into a performance guarantee for the compressive matched filter. This is codified in the following corollary.

Corollary 6. Suppose there exist constants $\alpha_1 \in [0, 1)$ and $\alpha_2 > 0$ such that $|R_{ss}(\tau)| \leq \alpha_1 R_{ss}(0)$ for all $|\tau| > \alpha_2$. Suppose also that $|\Omega||T| \geq 3$, that (19) is satisfied, and that

$$m > C_5 \cdot \max \left(\frac{\log(12|\Omega||T|/\delta)}{(1-\alpha_1)^2} \cdot \mu_1^2, \frac{\sqrt{\log(4/\delta)} \log(12|\Omega||T|/\delta)}{1-\alpha_1} \cdot \mu_2, \frac{\max(\log(|\Omega||T|), \log(2/\delta))}{(1-\alpha_1)^2} \cdot \frac{\sigma_n^2 |\Omega|}{|A|^2 \|\hat{s}\|_2^2} \right), \quad (21)$$

where C_5 is a known universal constant. Then with probability at least $1 - 2\delta$, the maximum value of $|\tilde{R}_{ss,n}(\tau)|$ must be attained for some $\tilde{\tau}_0$ within the interval $[\tau_0 - \alpha_2, \tau_0 + \alpha_2]$.

Proof. Using (16), we have that

$$\sup_{\tau} |\tilde{R}_{ss,n}(\tau) - AR_{ss}(\tau - \tau_0)| \leq \sup_{\tau} |\tilde{R}_{ss}(\tau) - AR_{ss}(\tau - \tau_0)| + \sup_{\tau} |\tilde{N}(\tau)|,$$

where $\tilde{R}_{ss}(\tau)$ is defined in (7). With probability at least $1 - 2\delta$, both (11) and (20) will be satisfied, and so we will have

$$|\tilde{R}_{ss,n}(\tau_0)| \geq |A| \frac{\|\hat{s}\|_2^2}{2\pi} - U - \frac{C_4 \sigma_n \sqrt{|\Omega|} \|\hat{s}\|_2 \max \left(\sqrt{\log(|\Omega||T|)}, \sqrt{\log(2/\delta)} \right)}{\sqrt{m}} \quad (22)$$

and

$$|\tilde{R}_{ss,n}(\tau)| \leq \alpha_1 |A| \frac{\|\hat{s}\|_2^2}{2\pi} + U + \frac{C_4 \sigma_n \sqrt{|\Omega|} \|\hat{s}\|_2 \max \left(\sqrt{\log(|\Omega||T|)}, \sqrt{\log(2/\delta)} \right)}{\sqrt{m}} \quad (23)$$

for all τ such that $|\tau - \tau_0| > \alpha_2$. If (21) is satisfied with $C_5 = \max(16C_1^2, 4C_1, 64\pi^2 C_4^2)$, then it follows that the right hand side of (22) must exceed the right hand side of (23). As with Corollary 3, one can slightly strengthen this result by choosing differing constants for the three terms appearing in (21). \square

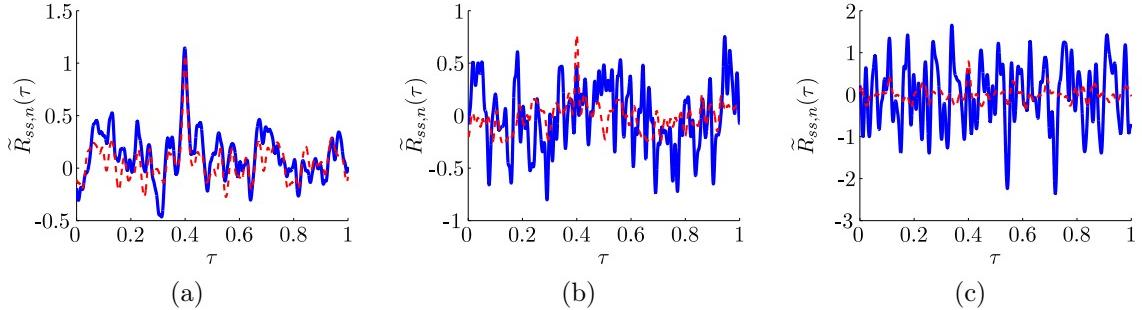


Figure 3: Estimated scaled, shifted autocorrelation function $\tilde{R}_{ss,n}(\tau)$ (solid blue line) obtained from noisy samples, and for the sake of comparison, the estimate $\tilde{R}_{ss}(\tau)$ (dashed red line) that would have been obtained without noise. For all experiments, the number of samples $m = 50$, and the noise level (a) $\sigma_n = 0.2 \cdot |A| \|\hat{s}\|_2 \sqrt{m/|\Omega|}$, (b) $\sigma_n = 0.5 \cdot |A| \|\hat{s}\|_2 \sqrt{m/|\Omega|}$, and (c) $\sigma_n = |A| \|\hat{s}\|_2 \sqrt{m/|\Omega|}$. Overall, the time-of-arrival estimation is reliable in the first case, tenuous in the second case, and completely unreliable in the third case.

To within a constant factor, the first two terms in (21) are the same as in Corollary 3; we might think of these terms as “activation” conditions for when the compressive matched filter will be well-behaved in the absence of noise. After these conditions are met, it can withstand noise levels up to a size

$$\sigma_n \sim |A| \|\hat{s}\|_2 \sqrt{\frac{m}{|\Omega| \log(|\Omega||T|)}}. \quad (24)$$

We can interpret (18) as the noise level at which the operation of the compressive matched filter will fall apart completely, and (24) (which is only a log factor smaller) as the noise level at which we have guaranteed accuracy.

Three examples of the estimated autocorrelation function $\tilde{R}_{ss,n}(\tau)$ are shown in Figure 3 for the Gaussian pulse example from Section 2.3 with a fixed number $m = 50$ of samples and various values of σ_n^2 (the noiseless estimates $\tilde{R}_{ss}(\tau)$ are overlaid). For the same number of measurements, Figure 4 shows the average performance of the compressive matched filter versus the noise level. For various noise levels σ_n between 0 and $|A| \|\hat{s}\|_2 \sqrt{m/|\Omega|}$, we run 1000 experiments generating random sample frequencies and random noise, and estimate τ_0 by identifying the peak of $|\tilde{R}_{ss,n}(\tau)|$. The figure indicates the percentage of trials in which the delay was estimated to within a distance $2a$ of the correct value τ_0 . We see in these experiments that the estimator begins to lose effectiveness roughly when $\sigma_n \sim 0.25 \cdot |A| \|\hat{s}\|_2 \sqrt{m/|\Omega|}$.

It is worth recalling that a user may have some control over selecting the observation interval Ω . In most cases it would be natural to choose Ω roughly equal to the essential bandwidth of s_0 . Taking Ω larger than this will increase μ_1 , μ_2 , and the sensitivity to noise; taking Ω smaller than this will generally increase the width of the main lobe of the autocorrelation function $R_{ss}(\tau)$ and thus limit the resolution to which τ_0 can be estimated.

Finally, we can compare the noise levels in (18) and (24) to the noise levels at which a digital matched filter working from a set of samples taken in the time domain at the Nyquist rate $|\Omega|/2\pi$ will stop being effective. Suppose we sample the (bandlimited) return signal

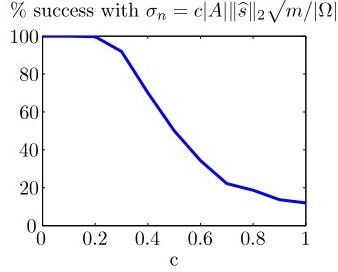


Figure 4: Percentage of correct time-of-arrival estimation ($|\hat{\tau}_0 - \tau_0| \leq 2a$) over 1000 trials, as a function of the noise level $\sigma_n = c \cdot |A| \|\hat{s}\|_2 \sqrt{m/|\Omega|}$, $0 \leq c \leq 1$.

$As(t - \tau_0)$ at the Nyquist rate, and noise is added to these samples. We observe

$$y_d[\ell] = As_{d,\tau_0}[\ell] + n_d[\ell], \quad \text{where } s_{d,\tau_0}[\ell] = s(t - \tau_0)|_{t=\ell 2\pi/|\Omega|},$$

and $n_d[\ell]$ is sequence of independent zero-mean Gaussian random variables with variances $\sigma_n^2 |\Omega| / 2\pi$. This variance is chosen to make the noise process similar to that analyzed in the compressive case; it corresponds to samples of a continuous-time process that has a power spectral density equal to σ_n^2 on Ω and zero elsewhere.

Focusing just on the complex case for the sake of brevity, once we have collected y_d , we can estimate the scaled, shifted autocorrelation function using

$$\tilde{R}_d(\tau) = \langle y_d, s_{d,\tau} \rangle = A \langle s_{d,\tau_0}, s_{d,\tau} \rangle + \langle n_d, s_{d,\tau} \rangle \quad (25)$$

and choose as our estimate of τ_0 the maximizer of $|\tilde{R}_d(\tau)|$ over all τ . At the correct shift τ_0 , the first inner product in (25) is given by

$$\langle s_{d,\tau_0}, s_{d,\tau_0} \rangle = \sum_{\ell} |s_{d,\tau_0}[\ell]|^2 = \frac{|\Omega|}{2\pi} \|s(t - \tau_0)\|_2^2 = \frac{|\Omega|}{4\pi^2} \|\hat{s}\|_2^2,$$

where the second equality comes from the fact that $s(\cdot)$ is bandlimited and we are sampling at the Nyquist rate. The second inner product is a Gaussian random variable with

$$\text{Var} [\langle n_d, s_{d,\tau} \rangle] = \frac{\sigma_n^2 |\Omega|}{2\pi} \sum_{\ell} |s_{d,\tau}[\ell]|^2 = \frac{\sigma_n^2 |\Omega|^2}{8\pi^3} \|\hat{s}\|_2^2,$$

and so

$$\mathbb{E} |\langle n_d, s_{d,\tau} \rangle| \leq \frac{\sigma_n |\Omega|}{2\pi \sqrt{2\pi}} \cdot \|\hat{s}\|_2.$$

Roughly speaking, then, the Nyquist sampled matched filter will be overwhelmed by the noise when

$$\sigma_n \sim |A| \|\hat{s}\|_2. \quad (26)$$

Comparing (26) to the compressive matched filter results (18) and (24), we can interpret the factor of $\sqrt{m/|\Omega|}$ as a sort of *undersampling* penalty; as the number of samples gets smaller, the noise tolerance gets worse. When $m \gtrsim |\Omega|$, the performance of the two schemes will be similar. (A similar undersampling penalty arises in standard CS [8], where the noise variance that can be tolerated for a given recovery error decreases as the number of measurements gets smaller.)

3 Pure Tone Estimation

As discussed in Section 1.2.3, the roles of time and frequency are completely interchangeable in our settings. This allows us to apply the compressive matched filter to the problem of finding the carrier frequency of a modulated signal from time-domain samples with minimal effort. One particular such application is studied here: estimating the frequency of a pure tone from random samples in time.

Formally the problem under study is described as follows. A pure exponential $Ae^{i\omega_0 t}$ with fixed—but unknown—frequency $\omega_0 \in \Omega$, amplitude $|A|$, and phase $\angle A$ is observed on $T = [-t_{\max}, t_{\max}]$. Let y be the vector of observations at sampling times $t_1, t_2, \dots, t_m \in T$, which are randomly chosen from a uniform distribution on T , i.e.,

$$y = A \begin{bmatrix} e^{i\omega_0 t_1} \\ e^{i\omega_0 t_2} \\ \vdots \\ e^{i\omega_0 t_m} \end{bmatrix}.$$

Given $y \in \mathbb{C}^m$, we are interested in estimating $\omega_0 \in \Omega$ and $A \in \mathbb{C}$. A natural approach to solving this problem is to find the find ω_0 and A which best explain the measurements in a least-squares sense. More formally, we define

$$(\widehat{\omega}_0, \widehat{A}) := \arg \min_{\omega, A} \sum_{k=1}^m |y[k] - A \cdot e^{i\omega t_k}|^2 = \arg \min_{\omega, A} \|y - A\psi_\omega\|_2^2, \quad (27)$$

where for any $\omega \in \Omega$, the test vector $\psi_\omega \in \mathbb{C}^m$ is given by:

$$\psi_\omega[k] = e^{i\omega t_k}, \quad k = 1, 2, \dots, m.$$

The least-squares solution for ω_0 is given by

$$\widehat{\omega}_0 = \arg \max_{\omega \in \Omega} |\langle y, \psi_\omega \rangle|, \quad (28)$$

and subsequent to estimating ω_0 , the least-squares estimate for A can be computed as $\widehat{A} = \langle y, \psi_{\widehat{\omega}_0} \rangle / \|\psi_{\widehat{\omega}_0}\|_2^{-2}$.

3.1 Analytical Framework

Equation (28) suggests a correlation-based strategy for estimating the unknown frequency ω_0 . In order to study the performance of such an estimator, let us define the random process $X(\omega) := \langle y, \psi_\omega \rangle$ on Ω , which has the mean function

$$\begin{aligned} \mathbb{E}X(\omega) &= A\mathbb{E} \sum_{k=1}^m e^{i(\omega_0-\omega)t_m} \\ &= A \sum_{k=1}^m \mathbb{E} e^{i(\omega_0-\omega)t_m} \\ &= mA|T|^{-1} \int_T e^{i(\omega_0-\omega)t} dt \\ &= mA|T|^{-1} \cdot |T| \text{sinc} \left(\frac{1}{2} |T| (\omega_0 - \omega) \right), \end{aligned}$$

where $\text{sinc}(\alpha) := \sin(\alpha)/\alpha$. One way to interpret $X(\omega)$ is that we have approximated the continuous-time inner product between two time-limited complex sinusoids as a discrete sum with samples taken at random locations; the above tells us that this estimate is unbiased.

Further attention reveals that we are facing the same problem as in the complex case of Section 2, where the roles of time and frequency have been interchanged: the frequency domain becomes the “shift domain,” while the time domain becomes the “observation domain.” More precisely, we may define $\widehat{s}_0(\omega) = 2\pi\delta(\omega)$ which has the inverse Fourier transform $s_0(t) = 1$. Our received signal can be expressed in the frequency domain as $A \cdot \widehat{s}_0(\omega - \omega_0)$ for some $\omega_0 \in \Omega$. However, we will observe m samples of this signal in the time domain, acquiring values of $Ae^{i\omega_0 t} s_0(t) = Ae^{i\omega_0 t}$ at times $t_1, t_2, \dots, t_m \in T$.

Now, in the observation (time) domain, we define $s(t)$ to be the time-limited version of $s_0(t)$, i.e., $s(t) := \mathbb{I}_{t \in T}$ where \mathbb{I} denotes the indicator function. Returning to the shift (frequency) domain, we have $\widehat{s}(\omega) = |T|\text{sinc}(\frac{1}{2}|T|\omega)$. Up to a constant factor, this expression equals its own autocorrelation function, i.e., $R_{\widehat{s}\widehat{s}}(\omega) = 2\pi\widehat{s}(\omega)$.

Therefore, we can estimate the ideal autocorrelation function $R_{\widehat{s}\widehat{s}}(\cdot)$ (up to the unknown complex amplitude A and translation ω_0) by rescaling the random process $X(\omega)$:

$$\widetilde{R}_{\widehat{s}\widehat{s}}(\omega) := \frac{2\pi|T|}{m} X(\omega). \quad (29)$$

This estimate is unbiased since $\mathbb{E}[\widetilde{R}_{\widehat{s}\widehat{s}}(\omega)] = A \cdot R_{\widehat{s}\widehat{s}}(\omega - \omega_0)$. It is clear that solving the least-squares problem (28) is equivalent to searching for the maximizer of $|\widetilde{R}_{\widehat{s}\widehat{s}}(\omega)|$. Since the main lobe of $R_{\widehat{s}\widehat{s}}(\omega) = 2\pi|T|\text{sinc}(\frac{1}{2}|T|\omega)$ is centered at the origin (with $R_{\widehat{s}\widehat{s}}(0) = 2\pi|T|$), we informally expect that, on average, finding the maximum of $|\widetilde{R}_{\widehat{s}\widehat{s}}(\omega)|$ correctly estimates ω_0 .

3.2 Noiseless Analysis

To study the concentration of $\widetilde{R}_{\widehat{s}\widehat{s}}(\omega)$ about its mean, we may follow the same arguments as in Sections 2.2 and 2.4 while simply exchanging the roles of time and frequency. In particular, we note that the problem of pure tone estimation corresponds to the first “special case” studied in Section 2.4, because the windowed signal template has uniform magnitude in the observation domain. Thus, we have $\mu_1 = \mu_2 = 1$. This leads us to the following result for the case of noiseless observations.

Corollary 7. *Fix $\delta > 0$ and let*

$$U = 2\pi C_1 |A| |T| \cdot \max \left(\frac{1}{\sqrt{m}}, \frac{\sqrt{\log(4/\delta)}}{m} \right) \cdot \sqrt{\log(12|\Omega||T|/\delta)}.$$

If $|\Omega||T| \geq 3$, then the estimate of the autocorrelation function in (29) obeys

$$\Pr \left\{ \sup_{\omega \in \Omega} \left| \widetilde{R}_{\widehat{s}\widehat{s}}(\omega) - A R_{\widehat{s}\widehat{s}}(\omega - \omega_0) \right| > U \right\} \leq \delta.$$

This corollary follows immediately from Theorem 2.

A close inspection of the definition of $\tilde{R}_{\hat{s}\hat{s}}(\omega)$ reveals that ω_0 is guaranteed to be a maximizer of $|\tilde{R}_{\hat{s}\hat{s}}(\omega)|$, with $|\tilde{R}_{\hat{s}\hat{s}}(\omega_0)| = 2\pi|A||T|$. What Corollary 7 ensures is that even for small values of m , no other values of ω far from ω_0 can give $|\tilde{R}_{\hat{s}\hat{s}}(\omega)|$ equal (or even close to) $|\tilde{R}_{\hat{s}\hat{s}}(\omega_0)|$. This fact is not only useful when we introduce nonidealities into the observation process (see Section 3.4) but also in guiding a computational method to search for the peak of $|\tilde{R}_{\hat{s}\hat{s}}(\omega)|$. We investigate this issue in Section 3.3 below.

3.3 A Grid Search Approach

In practice, in order to find the peak of $|\tilde{R}_{\hat{s}\hat{s}}(\omega)|$, one might hope to simply sample this function over a uniformly spaced grid of frequencies drawn from Ω . Because $\tilde{R}_{\hat{s}\hat{s}}(\omega)$ is guaranteed to remain close to $AR_{\hat{s}\hat{s}}(\omega - \omega_0)$, which decays sharply away from ω_0 , it is possible to guarantee that as long as the grid is chosen sufficiently fine, then the empirical maximum over the grid points will occur very close to the true peak.

To illustrate this fact with some specific but arbitrary values, let us note that for $|\omega| \leq \pi|T|^{-1}$, $|R_{\hat{s}\hat{s}}(\omega)| \geq 0.636 \cdot 2\pi|T|$. Moreover, for $|\omega| \geq 2\pi|T|^{-1}$, $|R_{\hat{s}\hat{s}}(\omega)| \leq 0.218 \cdot 2\pi|T|$. Following the techniques used to prove Corollary 3, we can ensure that $0.636 \cdot 2\pi|A||T| - U > 0.218 \cdot 2\pi|A||T| + U$ with probability at least $1 - \delta$ by taking

$$m \geq C_2 \cdot \max \left(\frac{\log(12|\Omega||T|/\delta)}{(0.636 - 0.218)^2}, \frac{\sqrt{\log(4/\delta) \log(12|\Omega||T|/\delta)}}{(0.636 - 0.218)} \right).$$

It follows that, if we initially search for the maximum of $|R_{\hat{s}\hat{s}}(\omega)|$ on a grid with resolution $2\pi|T|^{-1}$ (note that this is the so-called grid of Nyquist frequencies, given T), we are guaranteed that the empirical maximum will occur at a grid point $\hat{\omega}_0$ such that $|\hat{\omega}_0 - \omega_0| < 2\pi|T|^{-1}$.

After this initial grid search, it is actually straightforward to refine the accuracy of the estimate $\hat{\omega}_0$ using a local concave ascent. Note that

$$\begin{aligned} |\tilde{R}_{\hat{s}\hat{s}}(\omega)|^2 &= \left(\frac{2\pi|T|}{m} \right)^2 |\langle y, \psi_\omega \rangle|^2 \\ &= \left(\frac{2\pi|T||A|}{m} \right)^2 \sum_{i,j} e^{i(\omega_0 - \omega)(t_i - t_j)} \\ &= \frac{(2\pi|T||A|)^2}{m} + \left(\frac{2\pi|T||A|}{m} \right)^2 \sum_{i \neq j} \cos((\omega_0 - \omega)(t_i - t_j)). \end{aligned}$$

Since $|t_i - t_j| \leq |T|$, $|\tilde{R}_{\hat{s}\hat{s}}(\omega)|^2$ is guaranteed to be a concave function of ω when $|\omega - \omega_0| \leq \frac{\pi}{2} \cdot |T|^{-1}$. Therefore, if we have an estimate $\hat{\omega}_0$ sufficiently close to the true ω_0 (i.e., $|\hat{\omega}_0 - \omega_0| \leq \frac{\pi}{2} \cdot |T|^{-1}$), a standard concave maximization (akin to convex minimization) procedure beginning at $|\tilde{R}_{\hat{s}\hat{s}}(\hat{\omega}_0)|$ will give us the *exact* value for ω_0 . Since the grid search above guarantees that $|\hat{\omega}_0 - \omega_0| < 2\pi|T|^{-1}$, one could ensure success by running four concave maximizations starting from the points $\hat{\omega}_0 \pm \frac{\pi}{2} \cdot |T|^{-1}$ and $\hat{\omega}_0 \pm \frac{3\pi}{2} \cdot |T|^{-1}$.

3.4 Robustness

It is also possible to consider nonidealities in the observation process. Following the same set of arguments as in Section 2.5 (but exchanging the roles of time and frequency), we

arrive at the following result.

Corollary 8. *Let $N(\omega)$ denote the random process induced by additive complex-valued Gaussian measurement noise having variance σ_n^2 , and define*

$$\tilde{R}_{\hat{s}\hat{s},n}(\omega) := \frac{2\pi|T|}{m}(X(\omega) + N(\omega))$$

to be the estimate of the autocorrelation function formed using the noisy samples. Let $\delta > 0$. Suppose that $|\Omega||T| \geq 3$, that $m \geq C_3 \log(1/\delta)$, and that

$$m \geq C_5 \cdot \max \left(\frac{\log(12|\Omega||T|/\delta)}{(1 - 0.218)^2}, \frac{\sqrt{\log(4/\delta) \log(12|\Omega||T|/\delta)}}{1 - 0.218}, \frac{\max(\log(|\Omega||T|), \log(2/\delta))}{(1 - 0.218)^2} \cdot \frac{\sigma_n^2}{|A|^2} \right).$$

Then with probability at least $1 - 2\delta$, the maximum value of $|\tilde{R}_{\hat{s}\hat{s},n}(\omega)|$ must be attained for some $\hat{\omega}_0$ within the interval $[\omega_0 - 2\pi|T|^{-1}, \omega_0 + 2\pi|T|^{-1}]$.

Finally, let us note that with some additional work, we believe it would be possible to extend our analysis to account for multiple tones (or multiple translated pulses in the context of Section 2). The problem becomes that of detecting the true peaks in a noisy sum of sinc functions. For tones that are well-separated, one could argue that the interference in the random process is minimal and that any prominent peak in $|\tilde{R}_{\hat{s}\hat{s},n}(\omega)|$ indicates the presence of a tone. Tones that are very close may be impossible to discriminate (this is true even with Nyquist-rate samples), while tones that are moderately separated may be possible to discriminate by employing a greedy, iterative estimation procedure.

3.5 Stylized Application: Chirp Time-of-Arrival Estimation

We close by noting that the ability to estimate a pure tone’s frequency from random time samples can also be parlayed into a technique for estimating a chirp signal’s time-of-arrival from random time samples. For this discussion, suppose we receive a chirp signal

$$x(t) = A \exp \left(j \left(\omega_c(t - t_0) + \frac{\alpha}{2}(t - t_0)^2 \right) \right)$$

over some time interval, where ω_c denotes the known starting frequency, α denotes the known chirp rate, A denotes the complex amplitude, and t_0 denotes the unknown time-of-arrival. We can “de-chirp” this signal over this interval, computing

$$\tilde{x}(t) = x(t) \exp \left(-j \left(\omega_c t + \frac{\alpha}{2} t^2 \right) \right) = \tilde{A} e^{-j\alpha t_0 t},$$

where \tilde{A} is a complex amplitude. The signal $\tilde{x}(t)$ is merely a complex sinusoid (in this case, with frequency αt_0). We have argued in this section that it is possible to estimate a pure tone’s frequency from random samples in time, and in this case that means that it is possible to estimate the time-of-arrival parameter t_0 from random samples of $\tilde{x}(t)$ in time. It is important to note that time samples of $\tilde{x}(t)$ can be computed easily from time samples of $x(t)$ itself, since the two signals are related via point-wise multiplication.

4 Theory

4.1 Proofs of Theorems 1 and 2 (Noiseless Analysis)

Let us begin by noting that, in the real case, both Theorems 1 and 2 are concerned with bounding

$$\begin{aligned}
|\tilde{R}_{ss}(\tau) - AR_{ss}(\tau - \tau_0)| &= |\tilde{R}_{ss}(\tau) - A \cdot \operatorname{Re}(R_{ss}(\tau - \tau_0))| \\
&= \left| \frac{|\Omega|}{2\pi m} \cdot \operatorname{Re}(X(\tau)) - \frac{|\Omega|}{2\pi m} \cdot \operatorname{Re}(\mathbb{E}X(\tau)) \right| \\
&= \frac{|\Omega|}{2\pi m} |\operatorname{Re}(X(\tau) - \mathbb{E}X(\tau))| \\
&\leq \frac{|\Omega|}{2\pi m} |X(\tau) - \mathbb{E}X(\tau)|.
\end{aligned} \tag{30a}$$

In the complex case, both theorems are concerned with bounding

$$|\tilde{R}_{ss}(\tau) - AR_{ss}(\tau - \tau_0)| = \frac{|\Omega|}{2\pi m} |X(\tau) - \mathbb{E}X(\tau)|. \tag{30b}$$

Thus, to cover both cases, it suffices to focus on bounding $|X(\tau) - \mathbb{E}X(\tau)|$.

4.1.1 Setup

The first step in our approach to bounding $|X(\tau) - \mathbb{E}X(\tau)|$ is to define the centered process

$$Y(\tau) := X(\tau) - \mathbb{E}X(\tau) = A \sum_{k=1}^m |\hat{s}(\omega_k)|^2 e^{i\omega_k(\tau-\tau_0)} - 2\pi Am|\Omega|^{-1}R_{ss}(\tau - \tau_0).$$

Our goal is to bound $\sup_\tau |Y(\tau)|$, but to do this, we relate the random process to one that is more easily bounded. First, we symmetrize $Y(\tau)$ in the standard way. Create an independent copy $Y'(\tau)$ (generated from an independent set of samples $\omega'_1, \omega'_2, \dots, \omega'_m$), and define

$$\begin{aligned}
Z(\tau) &:= Y(\tau) - Y'(\tau) \\
&= A \sum_{k=1}^m |\hat{s}(\omega_k)|^2 e^{i\omega_k(\tau-\tau_0)} - |\hat{s}(\omega'_k)|^2 e^{i\omega'_k(\tau-\tau_0)}.
\end{aligned} \tag{31}$$

Each term in (31) is a symmetric random variable, and so $Z(\tau)$ has the same distribution as

$$Z'(\tau) := A \sum_{k=1}^m \epsilon_k \left(|\hat{s}(\omega_k)|^2 e^{i\omega_k(\tau-\tau_0)} - |\hat{s}(\omega'_k)|^2 e^{i\omega'_k(\tau-\tau_0)} \right),$$

where $\epsilon_1, \epsilon_2, \dots, \epsilon_m$ is a Rademacher sequence independent of everything.³

We can control $\mathbb{E} \sup_\tau |Y(\tau)|$ through $\mathbb{E} \sup_\tau |Z'(\tau)|$ using the following simple result, which is proved in Appendix A.

³A Rademacher sequence is a sequence of independent random variables taking ± 1 values with equal probabilities.

Lemma 9. $\mathbb{E} \sup_{\tau} |Y(\tau)| \leq \mathbb{E} \sup_{\tau} |Z'(\tau)|$.

Furthermore, the deviation of $\sup_{\tau} |Y(\tau)|$ from its average can be controlled through the corresponding deviation of $\sup_{\tau} |Z'(\tau)|$. The following is proved in Appendix B.

Lemma 10. For any $\lambda \geq 0$,

$$\mathbb{P} \left\{ \sup_{\tau} |Y(\tau)| > 2\mathbb{E} \sup_{\tau} |Y(\tau)| + \lambda \right\} \leq 2 \mathbb{P} \left\{ \sup_{\tau} |Z'(\tau)| > \lambda \right\}.$$

The above results allow us to focus on developing expectation and tail bounds for $\sup_{\tau} |Z'(\tau)|$. We establish such bounds in the following subsections.

To ease the notation below, we make the following definitions for quantities that will appear often:

$$\begin{aligned} M &= M(s) := \sup_{\omega \in \Omega} |\widehat{s}(\omega)|^2 = \|\widehat{s}\|_{\infty}^2, \\ M_1 &= M_1(s, m, \Omega) := \sqrt{\frac{m}{|\Omega|}} \|\widehat{s}\|_4^2, \\ M_2 &= M_2(s, m, \Omega) := \sqrt{\frac{m}{|\Omega|}} \|\widehat{s}\|_2, \\ M_3 &= M_3(s, \omega_1, \omega_2, \dots, \omega_m, \omega'_1, \omega'_2, \dots, \omega'_m) := \sqrt{\sum_{k=1}^m |\widehat{s}(\omega_k)|^4 + |\widehat{s}(\omega'_k)|^4}, \end{aligned}$$

and

$$M_4 = M_4(s, \omega_1, \omega_2, \dots, \omega_m) := \sqrt{\sum_{k=1}^m |\widehat{s}(\omega_k)|^2}.$$

We will also frequently use the following convenient facts. For any a and b , we have

$$e^{ia} - e^{ib} = 2i \sin\left(\frac{a-b}{2}\right) e^{i(a+b)/2}, \quad (32)$$

and

$$|a+b|^2 \leq 2|a|^2 + 2|b|^2. \quad (33)$$

Also, for any $c, u > 0$, the following inequality follows from a standard Gaussian tail bound [19]:

$$\int_{x \geq u} e^{-\frac{x^2}{c^2}} dx \leq \frac{c^2}{2u} e^{-\frac{u^2}{c^2}}. \quad (34)$$

4.1.2 Chaining

We start by bounding $\sup_{\tau} |Z'(\tau)|$ conditioned on the choice of $\{\omega_k\}$ and $\{\omega'_k\}$. To this end, we will use a chaining argument similar to what is used to prove the general Dudley inequality [12], but optimized for our particular process (this will allow us to tightly control the constants).

The following tail bounds for $Z'(\tau)$ and its increments are proved in Appendix C.

Lemma 11. *For a fixed $\tau \in \mathbb{R}$ and any $\lambda \geq 0$, $Z'(\tau)$ obeys*

$$P_{\epsilon_k} \{ |Z'(\tau)| > \lambda \} \leq 2 \exp \left(-\frac{\lambda^2}{4|A|^2 M_3^2} \right), \quad (35)$$

where P_{ϵ_k} denotes probability with respect to $\{\epsilon_k\}$ conditioned on fixed $\{\omega_k\}$ and $\{\omega'_k\}$. Also, for fixed $\tau_1, \tau_2 \in \mathbb{R}$,

$$P_{\epsilon_k} \{ |Z'(\tau_1) - Z'(\tau_2)| > \lambda \} \leq 2 \exp \left(-\frac{\lambda^2}{|A|^2 M_3^2 |\Omega|^2 |\tau_1 - \tau_2|^2} \right). \quad (36)$$

We will consider the values of $Z'(\tau)$ on a series of discrete grids of points that are essentially localized on the interval $T = [\tau_{\min}, \tau_{\max}]$. For each integer $j \geq 0$, let T_j be a grid of points spaced $2^{-j}|\Omega|^{-1}$ apart:

$$T_j = \{ \tau_{\min} + 2^{-j-1}|\Omega|^{-1} + k2^{-j}|\Omega|^{-1}, k = 0, 1, \dots, [2^j|\Omega||T|] \}. \quad (37)$$

All points in T_j belong to T , except possibly the final point in T_j , which may exceed τ_{\max} by no more than $2^{-j-1}|\Omega|^{-1}$. Moreover, if we denote by $\pi_j(\tau)$ the closest point in T_j to a given point τ , then $|\tau - \pi_j(\tau)| \leq 2^{-j-1}|\Omega|^{-1}$ for all $\tau \in T$. The points in the T_j are arranged like nodes in a dyadic tree, with each “parent” in T_j having two “children” in T_{j+1} (the two points that are closer to the parent than to any other point in T_j); the only exception to this rule occurs if $|T_{j+1}|$ is odd, in which case the final point in T_j has only one child in T_{j+1} .

We define L_j to be the set of “links” that connect the parents in T_j to their children in T_{j+1} :

$$L_j = \{ (p, q) \in (T_j, T_{j+1}) \mid \pi_j(\tau) = p \text{ and } \pi_{j+1}(\tau) = q \text{ for some } \tau \in T \}. \quad (38)$$

Because of the one-dimensional structure of T and the particular arrangement of T_j 's, we observe that every child in T_{j+1} is associated with only one link, and thus $\#L_j = \#T_{j+1} \leq 2^{j+1}|\Omega||T| + 1$. Furthermore, the length of every link is half of the distance between consecutive points on T_{j+1} ; that is $|q_j - p_j| = 2^{-j-2}|\Omega|^{-1}$ for all $(p_j, q_j) \in L_j$.

For almost every $\tau \in T$ [29, (6.46)], we can decompose $Z'(\tau)$ as a sum of the differences between approximations at different scales, writing the telescoping sum

$$Z'(\tau) = Z'(\pi_0(\tau)) + \sum_{j \geq 0} Z'(\pi_{j+1}(\tau)) - Z'(\pi_j(\tau)).$$

Thus

$$|Z'(\tau)| \leq |Z'(\pi_0(\tau))| + \sum_{j \geq 0} |Z'(\pi_{j+1}(\tau)) - Z'(\pi_j(\tau))|,$$

and

$$\sup_{\tau \in T} |Z'(\tau)| \leq \max_{p_0 \in T_0} |Z'(p_0)| + \sum_{j \geq 0} \max_{(p_j, q_j) \in L_j} |Z'(q_j) - Z'(p_j)|.$$

Therefore, for any $\lambda_1, \lambda_2 \geq 0$ and any sequence of positive numbers $\{u_j\}$ such that $\sum_{j \geq 0} u_j \leq$

1, we have

$$\begin{aligned}
& \mathbb{P}_{\epsilon_k} \left\{ \sup_{\tau \in T} |Z'(\tau)| > \lambda_1 + \lambda_2 \right\} \\
& \leq \mathbb{P}_{\epsilon_k} \left\{ \left\{ \max_{p_0 \in T_0} |Z'(\tau)| > \lambda_1 \right\} \cup \left\{ \sum_{j \geq 0} \max_{(p_j, q_j) \in L_j} |Z'(q_j) - Z'(p_j)| > \lambda_2 \right\} \right\} \\
& \leq \mathbb{P}_{\epsilon_k} \left\{ \max_{p_0 \in T_0} |Z'(\tau)| > \lambda_1 \right\} + \mathbb{P}_{\epsilon_k} \left\{ \sum_{j \geq 0} \max_{(p_j, q_j) \in L_j} |Z'(q_j) - Z'(p_j)| > \lambda_2 \right\} \\
& \leq \mathbb{P}_{\epsilon_k} \left\{ \max_{p_0 \in T_0} |Z'(\tau)| > \lambda_1 \right\} + \sum_{j \geq 0} \mathbb{P}_{\epsilon_k} \left\{ \max_{(p_j, q_j) \in L_j} |Z'(q_j) - Z'(p_j)| > \lambda_2 u_j \right\}. \quad (39)
\end{aligned}$$

To bound the first term in (39), we apply (35) along with the union bound and the fact that $\#T_0 \leq |\Omega||T| + 1$ to obtain

$$\mathbb{P}_{\epsilon_k} \left\{ \max_{p_0 \in T_0} |Z'(\tau)| > \lambda_1 \right\} \leq 2(\#T_0) \exp \left(\frac{-\lambda_1^2}{4|A|^2 M_3^2} \right) \leq 2(|\Omega||T| + 1) \exp \left(\frac{-\lambda_1^2}{4|A|^2 M_3^2} \right).$$

To bound the second term in (39), take $u_j = \sqrt{j+3} 2^{-j-2}$ and assume that $\lambda_2 \geq \lambda_0/3$, where $\lambda_0 := 3|A|M_3 \sqrt{\log(2|\Omega||T|)}$. Then, for every $j \geq 0$, we have

$$\begin{aligned}
\mathbb{P}_{\epsilon_k} \left\{ \max_{(p_j, q_j) \in L_j} |Z'(q_j) - Z'(p_j)| > \lambda_2 u_j \right\} & \leq 2(\#L_j) \exp \left(\frac{-\lambda_2^2 u_j^2}{|A|^2 M_3^2 |\Omega|^2 (2^{-j-2} |\Omega|^{-1})^2} \right) \\
& = 2(\#L_j) \exp \left(\frac{-(j+2)\lambda_2^2}{|A|^2 M_3^2} \right) \exp \left(\frac{-\lambda_2^2}{|A|^2 M_3^2} \right) \\
& \leq 2(\#L_j) (2|\Omega||T|)^{-j-2} \exp \left(\frac{-\lambda_2^2}{|A|^2 M_3^2} \right) \\
& \leq ((|\Omega||T|)^{-j-1} + 2^{-j-1} (|\Omega||T|)^{-j-2}) \exp \left(\frac{-\lambda_2^2}{|A|^2 M_3^2} \right),
\end{aligned}$$

where the first line above follows from applying (36) along with the union bound and the fact that $|q_j - p_j| = 2^{-j-2} |\Omega|^{-1}$ for all $(p_j, q_j) \in L_j$, the third line uses the assumption that $\lambda_2 \geq \lambda_0/3$, and the fourth line follows because $\#L_j \leq 2^{j+1} |\Omega||T| + 1$. If we assume that $|\Omega||T| \geq 3$, it follows that

$$\begin{aligned}
\sum_{j \geq 0} \mathbb{P}_{\epsilon_k} \left\{ \max_{(p_j, q_j) \in L_j} |Z'(q_j) - Z'(p_j)| > \lambda_2 u_j \right\} & \leq \sum_{j \geq 0} (3^{-j-1} + 2^{-j-1} \cdot 3^{-j-2}) \exp \left(\frac{-\lambda_2^2}{|A|^2 M_3^2} \right) \\
& \leq (17/30) \exp \left(\frac{-\lambda_2^2}{|A|^2 M_3^2} \right).
\end{aligned}$$

Putting together our bounds for the first and second terms in (39), for any $\lambda \geq \lambda_0$ we may

take $\lambda_1 = 2\lambda/3$ and $\lambda_2 = \lambda/3$ to conclude that

$$\begin{aligned} P_{\epsilon_k} \left\{ \sup_{\tau \in T} |Z'(\tau)| > \lambda \right\} &\leq (2|\Omega||T| + 2) \exp \left(\frac{-\lambda^2}{9|A|^2 M_3^2} \right) + (17/30) \exp \left(\frac{-\lambda^2}{9|A|^2 M_3^2} \right) \\ &\leq (2|\Omega||T| + 2.57) \exp \left(\frac{-\lambda^2}{9|A|^2 M_3^2} \right) \\ &\leq 3|\Omega||T| \exp \left(\frac{-\lambda^2}{9|A|^2 M_3^2} \right). \end{aligned} \quad (40)$$

The third line above follows from our assumption that $|\Omega||T| \geq 3$. In the subsections that follow, we translate this conditional tail bound into unconditional expectation and tail bounds for $\sup_{\tau} |Z'(\tau)|$.

4.1.3 Completing the Proof of Theorem 1 (Expectation)

Conditioned on the choice of $\{\omega_k\}$ and $\{\omega'_k\}$, we can integrate the tail bound developed above to obtain an upper bound for $E_{\epsilon_k} \sup_{\tau} |Z'(\tau)|$. Note that, for any nonnegative random variable V , we have [29, Prop. 6.1]

$$EV = \int_0^\infty P\{V > u\} du. \quad (41)$$

Once we have bounded the average of the supremum of the conditioned process, it is then straightforward to extend this to a bound for $E \sup_{\tau} |Z'(\tau)|$ by removing the conditioning on $\{\omega_k\}$ and $\{\omega'_k\}$.

Recall that $\lambda_0 = 3|A|M_3\sqrt{\log(2|\Omega||T|)}$. Then it follows from the identity above that

$$\begin{aligned} E_{\epsilon_k} \sup_{\tau \in T} |Z'(\tau)| &= \int_0^\infty P_{\epsilon_k} \left\{ \sup_{\tau \in T} |Z'(\tau)| > \lambda \right\} d\lambda \\ &= \int_0^{\lambda_0} P_{\epsilon_k} \left\{ \sup_{\tau \in T} |Z'(\tau)| > \lambda \right\} d\lambda + \int_{\lambda_0}^\infty P_{\epsilon_k} \left\{ \sup_{\tau \in T} |Z'(\tau)| > \lambda \right\} d\lambda \\ &\leq \lambda_0 + (2|\Omega||T| + 2.57) \int_{\lambda_0}^\infty \exp \left(\frac{-\lambda^2}{9|A|^2 M_3^2} \right) d\lambda \\ &\leq \lambda_0 + (2|\Omega||T| + 2.57) \cdot \frac{9|A|^2 M_3^2}{2\lambda_0} \exp \left(\frac{-\lambda_0^2}{9|A|^2 M_3^2} \right) \\ &= 3|A|M_3\sqrt{\log(2|\Omega||T|)} + \left(1 + \frac{1.285}{|\Omega||T|} \right) \frac{3|A|M_3}{2\sqrt{\log(2|\Omega||T|)}} \\ &\leq 3|A|M_3\sqrt{\log(2|\Omega||T|)} + \left(1 + \frac{1.285}{3} \right) \frac{3|A|M_3}{2\sqrt{\log 6}} \\ &\leq |A|M_3 \left(3\sqrt{\log(2|\Omega||T|)} + 1.61 \right). \end{aligned} \quad (42)$$

The fourth line above follows from (34), and the sixth line follows from our assumption that $|\Omega||T| \geq 3$. Now it remains to remove the conditioning by taking the average over $\{\omega_k\}$ and

$\{\omega'_k\}$. First note that, by Jensen's inequality, we have

$$\begin{aligned}
\mathbb{E}_{\omega_k, \omega'_k} M_3 &\leq \sqrt{\mathbb{E}_{\omega_k, \omega'_k} M_3^2} \\
&= \sqrt{\mathbb{E}_{\omega_k} \sum_{k=1}^m |\widehat{s}(\omega_k)|^4 + \mathbb{E}_{\omega'_k} \sum_{k=1}^m |\widehat{s}(\omega'_k)|^4} \\
&= \sqrt{2m\mathbb{E}_{\omega} |\widehat{s}(\omega)|^4} \\
&= \sqrt{2m|\Omega|^{-1} \|\widehat{s}\|_4^4} \\
&= \sqrt{2}M_1.
\end{aligned} \tag{43}$$

Now combining the above inequality with (42) brings us to

$$\begin{aligned}
\mathbb{E} \sup_{\tau \in T} |Z'(\tau)| &= \mathbb{E}_{\omega_k, \omega'_k} \mathbb{E}_{\epsilon_k} \sup_{\tau \in T} |Z'(\tau)| \\
&\leq \mathbb{E}_{\omega_k, \omega'_k} |A|M_3 \left(3\sqrt{\log(2|\Omega||T|)} + 1.61 \right) \\
&\leq |A|M_1 \left(4.25\sqrt{\log(2|\Omega||T|)} + 2.28 \right).
\end{aligned}$$

The final link to the random process of interest is via Lemma 9:

$$\begin{aligned}
\mathbb{E} \sup_{\tau} |Y(\tau)| &\leq \mathbb{E} \sup_{\tau} |Z'(\tau)| \\
&\leq |A|M_1 \left(4.25\sqrt{\log(2|\Omega||T|)} + 2.28 \right)
\end{aligned} \tag{44}$$

$$\begin{aligned}
&\leq |A|M_1 \left(4.25\sqrt{\log(2|\Omega||T|)} + \frac{2.28}{\sqrt{\log 6}} \sqrt{\log(2|\Omega||T|)} \right) \\
&\leq 5.96|A|M_1 \sqrt{\log(2|\Omega||T|)},
\end{aligned} \tag{45}$$

where the third line follows because we assumed that $|\Omega||T| \geq 3$. This completes the proof of Theorem 1, after plugging (44) and (45) into (30a) or (30b).

4.1.4 Completing the Proof of Theorem 2 (Tail Bound)

Recall the tail bound obtained in (40). In this section, we remove the conditioning on $\{\omega_k\}$ and $\{\omega'_k\}$ to obtain a tail bound for the supremum of $|Y(\tau)|$ on T . From (43), recall that M_3^2 is a sum of independent bounded random variables, and thus it is closely concentrated about its average $\mathbb{E}M_3^2 = 2M_1^2$. To quantify this, we will use the classic Bernstein inequality, which is restated below for convenience.

Lemma 12. [1] Consider a sequence of independent zero-mean random variables V_1, V_2, \dots, V_m with $|V_k| \leq B$ for $k = 1, 2, \dots, m$. Then for any $\lambda \geq 0$, the following holds:

$$\mathbb{P} \left\{ \sum_{k=1}^m V_k > \lambda \right\} \leq \exp \left(\frac{-\lambda^2}{2\rho^2 + 2B\lambda/3} \right),$$

where $\rho^2 = \sum_{k=1}^m \mathbb{E}V_k^2$.

Here, take $V_k = |\widehat{s}(\omega_k)|^4 + |\widehat{s}(\omega'_k)|^4 - 2|\Omega|^{-1} \|\widehat{s}\|_4^4$ and note that

$$\mathbb{E}V_k = 2\mathbb{E}\left(|\widehat{s}(\omega_k)|^4 - |\Omega|^{-1} \|\widehat{s}\|_4^4\right) = 0.$$

Also,

$$\begin{aligned} |V_k| &\leq \sup_{\omega, \omega'} \left| |\widehat{s}(\omega)|^4 + |\widehat{s}(\omega')|^4 - 2|\Omega|^{-1} \|\widehat{s}\|_4^4 \right| \\ &\leq \sup_{\omega, \omega'} \max \left(|\widehat{s}(\omega)|^4 + |\widehat{s}(\omega')|^4, 2|\Omega|^{-1} \|\widehat{s}\|_4^4 \right) \\ &= \max \left(2 \sup_{\omega} |\widehat{s}(\omega)|^4, 2|\Omega|^{-1} \|\widehat{s}\|_4^4 \right) \\ &= 2 \max \left(M^2, |\Omega|^{-1} \|\widehat{s}\|_4^4 \right) \\ &= 2M^2. \end{aligned}$$

The second line above follows from the convenient fact that $|a - b| \leq \max(a, b)$ for any $a, b \geq 0$, and the last line follows from the fact that $\|\widehat{s}\|_4^4 \leq |\Omega| \sup_{\omega} |\widehat{s}(\omega)|^4$. We also have

$$\begin{aligned} \rho^2 &= \sum_{k=1}^m \mathbb{E} \left(|\widehat{s}(\omega_k)|^4 + |\widehat{s}(\omega'_k)|^4 \right)^2 - 4|\Omega|^{-2} \|\widehat{s}\|_4^8 \\ &\leq \sum_{k=1}^m \mathbb{E} \left(|\widehat{s}(\omega_k)|^4 + |\widehat{s}(\omega'_k)|^4 \right)^2 \\ &\leq \sum_{k=1}^m 2\mathbb{E}|\widehat{s}(\omega_k)|^8 + 2\mathbb{E}|\widehat{s}(\omega'_k)|^8 \\ &= 4m\mathbb{E}|\widehat{s}(\omega)|^8 \\ &\leq 4mM^2\mathbb{E}|\widehat{s}(\omega)|^4 \\ &= 4mM^2|\Omega|^{-1}\|\widehat{s}\|_4^4 \\ &= 4M^2M_1^2. \end{aligned}$$

The first line above follows because $\mathbb{E}(V - \mathbb{E}V)^2 = \mathbb{E}V^2 - (\mathbb{E}V)^2$ for any real-valued random variable V . The third line is implied by (32). Now we can apply the Bernstein inequality to $\sum_{k=1}^m V_k$ for any $\lambda \geq 0$ and obtain

$$P\{M_3^2 > 2M_1^2 + \lambda\} \leq \exp\left(\frac{-\lambda^2}{8M^2M_1^2 + (4/3)M^2\lambda}\right). \quad (46)$$

Assume that $M_1 \geq M\sqrt{\log(4/\delta)}$. Then (46) implies that

$$P\{M_3^2 > 2M_1^2 + \lambda\} \leq \exp\left(\frac{-\lambda^2 \log(4/\delta)}{8M_1^4 + 4M_1^2\lambda/3}\right).$$

Take $\lambda = 3.58M_1^2$. Then

$$P\{M_3^2 > 5.58M_1^2\} \leq \delta/4. \quad (47)$$

Now assume that $M_1 \leq M\sqrt{\log(4/\delta)}$. Then

$$\begin{aligned} P\{M_3^2 > 2M^2 \log(4/\delta) + \lambda\} &\leq P\{M_3^2 > 2M_1^2 + \lambda\} \\ &\leq \exp\left(\frac{-\lambda^2}{8M^4 \log(4/\delta) + (4/3)M^2\lambda}\right). \end{aligned}$$

Take $\lambda = 3.58M^2 \log(4/\delta)$. Then

$$P\{M_3^2 > 5.58M^2 \log(4/\delta)\} \leq \delta/4. \quad (48)$$

Therefore, combining (47) and (48), we arrive at

$$P\left\{M_3 > 2.37 \max\left(M_1, M\sqrt{\log(4/\delta)}\right)\right\} \leq \delta/4.$$

Let \mathcal{E} denote the event that $M_3 \leq 2.37 \max\left(M_1, M\sqrt{\log(4/\delta)}\right)$. Then clearly,

$$P\{\mathcal{E}\} \geq 1 - \delta/4. \quad (49)$$

On the other hand, taking $\lambda = 3|A|M_3\sqrt{\log(12|\Omega||T|/\delta)}$, (40) implies that

$$P_{\epsilon_k}\left\{\sup_{\tau}|Z'(\tau)| > 3|A|M_3\sqrt{\log(12|\Omega||T|/\delta)}\right\} \leq \delta/4, \quad (50)$$

Now we can combine (49) and (50) as follows. For notational convenience, set $b := 3|A|M_3\sqrt{\log(12|\Omega||T|/\delta)}$, define

$$u := 7.11|A|\max\left(M_1, M\sqrt{\log(4/\delta)}\right)\sqrt{\log(12|\Omega||T|/\delta)},$$

and note that

$$\begin{aligned} & P\left\{\sup_{\tau \in T}|Z'(\tau)| > u\right\} \\ & \leq P\left\{\sup_{\tau \in T}|Z'(\tau)| > u \mid \mathcal{E}\right\} P\{\mathcal{E}\} + P\{\mathcal{E}^C\} \\ & \leq P\left\{\sup_{\tau \in T}|Z'(\tau)| > b \mid \mathcal{E}\right\} P\{\mathcal{E}\} + \frac{\delta}{4} \\ & = P\{\mathcal{E}\} \cdot \frac{1}{P\{\mathcal{E}\}} \int_{\mathcal{E}} P_{\epsilon_k}\left\{\sup_{\tau \in T}|Z'(\tau)| > b \mid \{\omega_k\}, \{\omega'_k\}\right\} d\mu(\{\omega_k\}, \{\omega'_k\}) + \frac{\delta}{4} \\ & \leq \int_{\mathcal{E}} \frac{\delta}{4} d\mu(\omega_1, \omega_2, \dots, \omega_m) + \frac{\delta}{4} \\ & = \frac{\delta}{2}. \end{aligned}$$

The fourth line above follows from the definition of conditional measure. The final link to the random process of interest is via Lemma 10:

$$\begin{aligned} P\left\{\sup_{\tau}|Y(\tau)| > 8.5|A|M_1\sqrt{\log(2|\Omega||T|)} + 4.56|A|M_1 + u\right\} & \leq P\left\{\sup_{\tau}|Y(\tau)| > 2\mathbb{E}\sup_{\tau}|Y(\tau)| + u\right\} \\ & \leq 2P\left\{\sup_{\tau}|Z'(\tau)| > u\right\} \\ & \leq \delta, \end{aligned} \quad (51)$$

where we used (44) in the first line above. This completes the proof of Theorem 2, as

$$\begin{aligned}
& 8.5|A|M_1\sqrt{\log(2|\Omega||T|)} + 4.56|A|M_1 + u \\
& \leq |A| \max(M_1, M\sqrt{\log(4/\delta)}) \left(8.5\sqrt{\log(2|\Omega||T|)} + 4.56 + 7.11\sqrt{\log(12|\Omega||T|/\delta)} \right) \\
& \leq |A| \max(M_1, M\sqrt{\log(4/\delta)}) \left(15.61\sqrt{\log(12|\Omega||T|/\delta)} + 4.56 \right) \\
& \leq |A| \max(M_1, M\sqrt{\log(4/\delta)}) \left(15.61\sqrt{\log(12|\Omega||T|/\delta)} + \frac{4.56}{\sqrt{\log 36}}\sqrt{\log(12|\Omega||T|/\delta)} \right) \\
& \leq C_1|A| \max(M_1, M\sqrt{\log(4/\delta)}) \sqrt{\log(12|\Omega||T|/\delta)},
\end{aligned}$$

where the fourth line follows from the assumption that $|\Omega||T| \geq 3$, and the fifth line holds by taking $C_1 = 18.02$. Actually, we see from the above that we may slightly improve upon the value of U specified in (10) by taking

$$U = \max\left(\frac{\eta\mu_1}{\sqrt{m}}, \frac{\eta\mu_2}{m} \cdot \sqrt{\log(4/\delta)}\right) \cdot \left(15.61\sqrt{\log(12|\Omega||T|/\delta)} + 4.56\right).$$

4.2 Proofs of Theorems 4 and 5 (Noisy Measurements)

Let us begin by noting that, in the real case, both Theorems 4 and 5 are concerned with bounding $|\tilde{N}(\tau)| = \frac{|\Omega|}{2\pi m} |\operatorname{Re}(N(\tau))| \leq \frac{|\Omega|}{2\pi m} |N(\tau)|$. In the complex case, both theorems are concerned with bounding $|\tilde{N}(\tau)| = \frac{|\Omega|}{2\pi m} |N(\tau)|$. Thus, to cover both cases, it suffices to focus on bounding $|N(\tau)|$.

We first bound $\mathbb{E} \sup_\tau |N(\tau)|$, and then we show that $\sup_\tau |N(\tau)|$ is sharply concentrated about its mean with high probability.

4.2.1 Proof of Theorem 4 (Expectation)

We begin by noting that conditioned on $\{\omega_k\}$, $N(\tau)$ is a complex-valued Gaussian process to which we can apply a chaining argument similar to the one put forth in Section 4.1. With the $\{\omega_k\}$ fixed, at each τ the real and imaginary parts of $N(\tau)$ have the same Gaussian distributions, and thus the magnitude $|N(\tau)|$ is a Rayleigh random variable with second moment

$$\begin{aligned}
\mathbb{E}_{n_k} |N(\tau)|^2 &= \mathbb{E}_{n_k} \left| \sum_{k=1}^m n_k \hat{s}^*(\omega_k) e^{i\omega_k \tau} \right|^2 \\
&= \sum_{k=1}^m \mathbb{E}_{n_k} |n_k|^2 |\hat{s}(\omega_k)|^2 \\
&= \sigma_n^2 \sum_{k=1}^m |\hat{s}(\omega_k)|^2 \\
&= \sigma_n^2 M_4^2,
\end{aligned} \tag{52}$$

where the second line follows from the independence of $\{n_k\}$. It is known that a Rayleigh random variable V with $\mathbb{E}V^2 = c^2$ satisfies $P\{V > \lambda\} = \exp(-\lambda^2/c^2)$ [19], and thus a tail

bound for $|N(\tau)|$ follows directly from the above:

$$\mathbb{P}_{n_k} \{|N(\tau)| > \lambda\} = \exp \left(\frac{-\lambda^2}{\sigma_n^2 M_4^2} \right). \quad (53)$$

Likewise, the increment $|N(\tau_1) - N(\tau_2)|$ is Rayleigh with

$$\begin{aligned} \mathbb{E}_{n_k} |N(\tau_1) - N(\tau_2)|^2 &= \mathbb{E}_{n_k} \left| \sum_{k=1}^m n_k \widehat{s}^*(\omega_k) (e^{i\omega_k \tau_1} - e^{i\omega_k \tau_2}) \right|^2 \\ &= \sum_{k=1}^m \mathbb{E}_{n_k} |n_k|^2 |\widehat{s}(\omega_k)|^2 |e^{i\omega_k \tau_1} - e^{i\omega_k \tau_2}|^2 \\ &= \sigma_n^2 \sum_{k=1}^m |\widehat{s}(\omega_k)|^2 |e^{i\omega_k \tau_1} - e^{i\omega_k \tau_2}|^2 \\ &= \sigma_n^2 \sum_{k=1}^m 4|\widehat{s}(\omega_k)|^2 \sin^2(\omega_k(\tau_1 - \tau_2)/2) \\ &\leq \sigma_n^2 \sum_{k=1}^m |\widehat{s}(\omega_k)|^2 |\omega_k|^2 |\tau_1 - \tau_2|^2 \\ &\leq \frac{1}{4} \sigma_n^2 |\Omega|^2 |\tau_1 - \tau_2|^2 \sum_{k=1}^m |\widehat{s}(\omega_k)|^2 \\ &= \frac{1}{4} \sigma_n^2 M_4^2 |\Omega|^2 |\tau_1 - \tau_2|^2. \end{aligned}$$

The second line above follows from the independence of $\{n_k\}$, and the fourth line follows from (32). The tail bound for the increment is then

$$\mathbb{P}_{n_k} (|N(\tau_1) - N(\tau_2)| > \lambda) \leq \exp \left(\frac{-4\lambda^2}{\sigma_n^2 M_4^2 |\Omega|^2 |\tau_1 - \tau_2|^2} \right). \quad (54)$$

With the same definition of the sets T_j and L_j from (37) and (38) in Section 4.1.2, we can write the telescoping sum for $N(\tau)$ and proceed similarly to obtain

$$\sup_{\tau \in T} |N(\tau)| \leq \max_{p_0 \in T_0} |N(p_0)| + \sum_{j \geq 0} \max_{(p_j, q_j) \in L_j} |N(q_j) - N(p_j)|,$$

and so it follows immediately that

$$\mathbb{E}_{n_k} \sup_{\tau \in T} |N(\tau)| \leq \mathbb{E}_{n_k} \max_{p_0 \in T_0} |N(p_0)| + \sum_{j \geq 0} \mathbb{E}_{n_k} \max_{(p_j, q_j) \in L_j} |N(q_j) - N(p_j)|. \quad (55)$$

We now use the following standard result that bounds the expected maximum of a finite set of subgaussian random variables. The proof is included in Appendix D.

Proposition 13. *Let V_1, V_2, \dots, V_N be random variables with $\mathbb{P}\{|V_i| > \lambda\} \leq K e^{-\lambda^2/2\sigma^2}$. Then*

$$\mathbb{E} \max_{1 \leq i \leq N} |V_i| \leq \sigma \left(\sqrt{2 \log(KN)} + \frac{1}{\sqrt{2 \log(KN)}} \right).$$

Applying Proposition 13 along with (53) allows us to bound the first term in (55):

$$\begin{aligned}
\mathbb{E}_{n_k} \max_{p_0 \in T_0} |N(p_0)| &\leq \frac{\sigma_n M_4}{\sqrt{2}} \left(\sqrt{2 \log(|\Omega||T| + 1)} + \frac{1}{\sqrt{2 \log(|\Omega||T|)}} \right) \\
&\leq \frac{\sigma_n M_4}{\sqrt{2}} \left(\sqrt{2 \log(|\Omega||T|)} + \frac{|\Omega||T| + 1}{|\Omega||T| \sqrt{2 \log(|\Omega||T|)}} \right) \\
&\leq \frac{\sigma_n M_4}{\sqrt{2}} \left(\sqrt{2 \log(|\Omega||T|)} + 0.9 \right) \\
&\leq \sigma_n M_4 \left(\sqrt{\log(|\Omega||T|)} + 0.64 \right).
\end{aligned}$$

The first line uses the fact that $|\Omega||T| \leq \#T_0 \leq |\Omega||T| + 1$, the second line follows from the convenient fact that for any $a > 1$ we have

$$\sqrt{\log(a+1)} - \sqrt{\log a} \leq \left(2a\sqrt{\log a} \right)^{-1}, \quad (56)$$

and the third line follows from the assumption that $|\Omega||T| \geq 3$.

For the multiscale sum of expected supremums in (55), we can apply Proposition 13 with $\sigma = \frac{1}{2\sqrt{2}}\sigma_n M_4 |\Omega||q_j - p_j| = \frac{2^{-j}}{8\sqrt{2}}\sigma_n M_4$ for every $j \geq 0$, and obtain

$$\begin{aligned}
\sum_{j \geq 0} \mathbb{E}_{n_k} \max_{(p_j, q_j) \in L_j} |N(p_j) - N(q_j)| &\leq \frac{\sigma_n M_4}{8\sqrt{2}} \sum_{j \geq 0} 2^{-j} \left(\sqrt{2 \log(2^{j+1}|\Omega||T| + 1)} + \frac{1}{\sqrt{2 \log(2^{j+1}|\Omega||T|)}} \right) \\
&\leq \frac{\sigma_n M_4}{8\sqrt{2}} \sum_{j \geq 0} 2^{-j} \left(\sqrt{2 \log(2^{j+1}|\Omega||T|)} + \frac{2^{j+1}|\Omega||T| + 1}{2^{j+1}|\Omega||T| \sqrt{2 \log(2^{j+1}|\Omega||T|)}} \right) \\
&\leq \frac{\sigma_n M_4}{8\sqrt{2}} \sum_{j \geq 0} 2^{-j} \left(\sqrt{2 \log(|\Omega||T|)} + \sqrt{2(j+1)\log 2} + 0.62 \right) \\
&\leq \frac{\sigma_n M_4}{8\sqrt{2}} \left(2\sqrt{2 \log(|\Omega||T|)} + 4.42 \right) \\
&\leq \sigma_n M_4 \left(0.25\sqrt{\log(|\Omega||T|)} + 0.4 \right),
\end{aligned}$$

where the first line uses the fact that $2^{j+1}|\Omega||T| \leq \#L_j \leq 2^{j+1}|\Omega||T| + 1$, the second line uses (56), and the third line uses the assumption that $|\Omega||T| \geq 3$. Consequently, in light of (55), we obtain

$$\mathbb{E}_{n_k} \sup_{\tau \in T} |N(\tau)| \leq \sigma_n M_4 \left(1.25\sqrt{\log(|\Omega||T|)} + 1.04 \right). \quad (57)$$

Removing the conditioning on $\{\omega_k\}$, we arrive at

$$\begin{aligned}
\mathbb{E} \sup_{\tau \in T} |N(\tau)| &= \mathbb{E}_{\omega_k} \mathbb{E}_{n_k} \sup_{\tau \in T} |N(\tau)| \\
&\leq \sigma_n \left(1.25 \sqrt{\log(|\Omega||T|)} + 1.04 \right) \mathbb{E}_{\omega_k} M_4 \\
&\leq \sigma_n \left(1.25 \sqrt{\log(|\Omega||T|)} + 1.04 \right) \sqrt{\sum_{k=1}^m \mathbb{E}_{\omega_k} |\hat{s}(\omega_k)|^2} \\
&= \sigma_n \left(1.25 \sqrt{\log(|\Omega||T|)} + 1.04 \right) \sqrt{m|\Omega|^{-1} \int_{\Omega} |\hat{s}(\omega)|^2 d\omega} \\
&= \sigma_n M_2 \left(1.25 \sqrt{\log(|\Omega||T|)} + 1.04 \right) \\
&\leq 2.25 \sigma_n M_2 \sqrt{\log(|\Omega||T|)},
\end{aligned}$$

where the third line uses Jensen's inequality, and the last line follows from our assumption that $|\Omega||T| \geq 3$. This completes the proof of Theorem 4.

4.2.2 Tail Bound (Proof of Theorem 5)

Recall that, conditioned on $\{\omega_k\}$, $N(\tau)$ is a centered complex-valued Gaussian process. The following result, proved in Appendix E, provides a sharp tail bound for the supremum of this random process.

Lemma 14. *Let $\{G(t), t \in \Delta\}$ be a centered complex-valued Gaussian process. Define the weak variance as $\nu^2 := \sup_{t \in \Delta} \mathbb{E} |G(t)|^2$. Then the following holds for any $\lambda \geq 0$:*

$$P \left\{ \sup_{t \in \Delta} |G(t)| \geq \mathbb{E} \sup_{t \in \Delta} |G(t)| + \lambda \right\} \leq \exp \left(-\frac{\lambda^2}{2\nu^2} \right).$$

To apply this bound to $N(\tau)$ conditioned on $\{\omega_k\}$, notice that (52) directly implies $\nu^2 = \sigma_n^2 M_4^2$. Therefore, using Lemma 14, we obtain the following for any $\lambda \geq 0$:

$$P_{n_k} \left\{ \sup_{\tau \in T} |N(\tau)| > \mathbb{E}_{n_k} \sup_{\tau \in T} |N(\tau)| + \lambda \right\} \leq \exp \left(-\frac{\lambda^2}{2\sigma_n^2 M_4^2} \right). \quad (58)$$

Now, recall (57) and note that due to our assumption that $|\Omega||T| \geq 3$, we have $\mathbb{E}_{n_k} \sup_{\tau \in T} |N(\tau)| \leq 2.25 \sigma_n M_4 \sqrt{\log(|\Omega||T|)}$. Combining this fact with (58), take $\lambda = \sigma_n M_4 \sqrt{2 \log(2/\delta)}$ to get

$$\begin{aligned}
&P_{n_k} \left\{ \sup_{\tau \in T} |N(\tau)| > 2\sigma_n M_4 \max \left(2.25 \sqrt{\log(|\Omega||T|)}, \sqrt{2 \log(2/\delta)} \right) \right\} \\
&\leq P_{n_k} \left\{ \sup_{\tau \in T} |N(\tau)| > 2.25 \sigma_n M_4 \sqrt{\log(|\Omega||T|)} + \sigma_n M_4 \sqrt{2 \log(2/\delta)} \right\} \\
&\leq \delta/2.
\end{aligned} \quad (59)$$

Now, we need only to show that M_4 is small with high probability. For this, we will use the Bernstein inequality (Lemma 12). To apply the Bernstein inequality, take $V_k =$

$|\widehat{s}(\omega_k)|^2 - |\Omega|^{-1} \|\widehat{s}\|_2^2$. Notice that

$$\mathbb{E}V_k = \mathbb{E}|\widehat{s}(\omega_k)|^2 - |\Omega|^{-1} \|\widehat{s}\|_2^2 = \int_{\Omega} |\widehat{s}(\omega)|^2 |\Omega|^{-1} d\omega - |\Omega|^{-1} \|\widehat{s}\|_2^2 = |\Omega|^{-1} \|\widehat{s}\|_2^2 - |\Omega|^{-1} \|\widehat{s}\|_2^2 = 0$$

and also that

$$\begin{aligned} |V_k| &\leq \sup_{\omega} \left| |\widehat{s}(\omega)|^2 - |\Omega|^{-1} \|\widehat{s}\|_2^2 \right| \\ &\leq \sup_{\omega} \max \left(|\widehat{s}(\omega)|^2, |\Omega|^{-1} \|\widehat{s}\|_2^2 \right) \\ &= \max \left(\sup_{\omega} |\widehat{s}(\omega)|^2, |\Omega|^{-1} \|\widehat{s}\|_2^2 \right) \\ &= \max \left(M, |\Omega|^{-1} \|\widehat{s}\|_2^2 \right) \\ &= M, \end{aligned}$$

where the second line uses the convenient fact that $|a - b| \leq \max(a, b)$ for any $a, b \geq 0$, and the fifth line uses the fact that $\|\widehat{s}\|_2^2 \leq |\Omega| \sup_{\omega} |\widehat{s}(\omega)|^2$. In addition, we know that

$$\mathbb{E}V_k^2 = \mathbb{E}|\widehat{s}(\omega_k)|^4 - |\Omega|^{-2} \|\widehat{s}\|_2^4 \leq \mathbb{E}|\widehat{s}(\omega_k)|^4 = |\Omega|^{-1} \|\widehat{s}\|_4^4,$$

where the first equality follows from the convenient fact that $\mathbb{E}|V - \mathbb{E}V|^2 = \mathbb{E}|V|^2 - |\mathbb{E}V|^2$ for any random variable V . Therefore, we have $\rho^2 = \sum_{k=1}^m \mathbb{E}V_k^2 \leq m|\Omega|^{-1} \|\widehat{s}\|_4^4 = M_1^2$. Now we can apply the Bernstein inequality to $\sum_{k=1}^m V_k$ for any $\lambda \geq 0$ and obtain

$$P \left\{ M_4^2 > m|\Omega|^{-1} \|\widehat{s}\|_2^2 + \lambda \right\} \leq \exp \left(-\frac{\lambda^2}{2M_1^2 + 2M\lambda/3} \right).$$

Suppose first that $M_1 \geq (M/3)\sqrt{\log(2/\delta)}$. Then take $\lambda = aM_1\sqrt{\log(2/\delta)}$ for some $a \geq 0$ to get

$$\begin{aligned} P \left\{ M_4^2 \geq m|\Omega|^{-1} \|\widehat{s}\|_2^2 + aM_1\sqrt{\log(2/\delta)} \right\} &\leq \exp \left(-\frac{a^2 M_1^2 \log(2/\delta)}{2M_1^2 + 2aMM_1\sqrt{\log(2/\delta)}/3} \right) \\ &\leq \exp \left(-\frac{a^2 M_1^2 \log(2/\delta)}{2M_1^2 + 2aM_1^2} \right) \\ &\leq \exp \left(-\frac{a^2}{2 + 2a} \log(2/\delta) \right) \\ &\leq \delta/2, \end{aligned}$$

which is valid for $a \geq 1 + \sqrt{3}$.

Now suppose that $M_1 < (M/3)\sqrt{\log(2/\delta)}$. Take $\lambda = aM \log(2/\delta)$. Then

$$\begin{aligned} P \left\{ M_4^2 \geq m|\Omega|^{-1} \|\widehat{s}\|_2^2 + aM \log(2/\delta) \right\} &\leq \exp \left(-\frac{a^2 M^2 \log^2(2/\delta)}{2M_1^2 + (2a/3)M^2 \log(2/\delta)} \right) \\ &< \exp \left(-\frac{a^2 M^2 \log^2(2/\delta)}{(2/9)M^2 \log(2/\delta) + (2a/3)M^2 \log(2/\delta)} \right) \\ &= \exp \left(-\frac{a^2 \log(2/\delta)}{2/9 + 2a/3} \right) \\ &\leq \delta/2, \end{aligned}$$

which is valid for $a \geq \frac{1+\sqrt{3}}{3}$.

So with probability at least $1 - \delta/2$ we will have

$$M_4^2 \leq m|\Omega|^{-1} \|\hat{s}\|_2^2 + a \max \left(M_1 \sqrt{\log(2/\delta)}, M \log(2/\delta) \right), \quad (60)$$

for $a \geq 1 + \sqrt{3}$.

By the definition of M_1 , note that $m|\Omega|^{-1} \|\hat{s}\|_2^2 \geq aM_1\sqrt{\log(2/\delta)}$ is equivalent to $m|\Omega|^{-1} \|\hat{s}\|_2^2 \geq a\sqrt{m}\sqrt{|\Omega|^{-1}} \|\hat{s}\|_4^2 \sqrt{\log(2/\delta)}$, which is equivalent to $m \geq a^2 \frac{\|\hat{s}\|_4^4}{\|\hat{s}\|_2^4} |\Omega| \log(2/\delta)$. Also, $m|\Omega|^{-1} \|\hat{s}\|_2^2 \geq aM \log(2/\delta)$ is equivalent to $m \geq a \frac{M}{\|\hat{s}\|_2^2} |\Omega| \log(2/\delta)$. Therefore, in order for the first term on the right hand side of (60) to be dominant we conveniently assume that

$$m \geq a^2 \max \left(\frac{|\Omega| \|\hat{s}\|_4^4}{\|\hat{s}\|_2^4}, \frac{|\Omega| M}{\|\hat{s}\|_2^2} \right) \log(2/\delta), \quad (61)$$

where we used the fact that $a > 1$. Under this assumption,

$$M_4^2 \leq 2m|\Omega|^{-1} \|\hat{s}\|_2^2 = 2M_2^2 \quad (62)$$

with probability exceeding $1 - \delta/2$. Let \mathcal{E} denote the event specified in (62); clearly if (61) is met, then $P\{\mathcal{E}\} \geq 1 - \delta/2$. Now, using (59), we have

$$\begin{aligned} & P \left\{ \sup_{\tau \in T} |N(\tau)| > 2\sqrt{2}\sigma_n M_2 \max \left(2.25\sqrt{\log(|\Omega||T|)}, \sqrt{2\log(2/\delta)} \right) \middle| \mathcal{E} \right\} \\ & \leq P \left\{ \sup_{\tau \in T} |N(\tau)| > 2\sigma_n M_4 \max \left(2.25\sqrt{\log(|\Omega||T|)}, \sqrt{2\log(2/\delta)} \right) \middle| \mathcal{E} \right\} \\ & = \frac{1}{P\{\mathcal{E}\}} \int_{\mathcal{E}} P_{n_k} \left\{ \sup_{\tau \in T} |N(\tau)| > 2\sigma_n M_4 \max \left(2.25\sqrt{\log(|\Omega||T|)}, \sqrt{2\log(2/\delta)} \right) \middle| \{\omega_k\} \right\} d\mu(\{\omega_k\}) \\ & \leq \frac{\delta}{2P\{\mathcal{E}\}} \int_{\mathcal{E}} d\mu(\{\omega_k\}) \\ & = \frac{\delta}{2}, \end{aligned} \quad (63)$$

where the third line above follows from the definition of conditional measure.

On the other hand, for any $u \geq 0$, we have

$$\begin{aligned} P \left\{ \sup_{\tau \in T} |N(\tau)| \geq u \right\} & = P \left\{ \sup_{\tau \in T} |N(\tau)| \geq u \middle| \mathcal{E} \right\} P\{\mathcal{E}\} + P \left\{ \sup_{\tau \in T} |N(\tau)| \geq u \middle| \mathcal{E}^C \right\} P\{\mathcal{E}^C\} \\ & \leq P \left\{ \sup_{\tau \in T} |N(\tau)| \geq u \middle| \mathcal{E} \right\} + \frac{\delta}{2}. \end{aligned}$$

Taking $u = 2\sqrt{2}\sigma_n M_2 \max \left(2.25\sqrt{\log(|\Omega||T|)}, \sqrt{2\log(2/\delta)} \right)$ and using (63), we finally obtain

$$\begin{aligned} & P \left\{ \sup_{\tau \in T} |N(\tau)| > 2\sqrt{2}\sigma_n M_2 \max \left(2.25\sqrt{\log(|\Omega||T|)}, \sqrt{2\log(2/\delta)} \right) \right\} \\ & \leq P \left\{ \sup_{\tau \in T} |N(\tau)| > 2\sqrt{2}\sigma_n M_2 \max \left(2.25\sqrt{\log(|\Omega||T|)}, \sqrt{2\log(2/\delta)} \right) \middle| \mathcal{E} \right\} + \frac{\delta}{2} \\ & \leq \delta, \end{aligned}$$

which is valid when (61) is met and $a \geq 1 + \sqrt{3}$. Setting $a = 1 + \sqrt{3}$, $C_3 = a^2$, and $C_4 = \frac{2.25\sqrt{2}}{\pi}$, we complete the proof of Theorem 5.

A Proof of Lemma 9

Let \mathbb{E}_{ω_k} and $\mathbb{E}_{\omega'_k}$ denote expectation with respect to $\{\omega_k\}$ and $\{\omega'_k\}$, respectively. Then we can write

$$\begin{aligned}
 \mathbb{E} \sup_{\tau} |Y(\tau)| &= \mathbb{E}_{\omega_k} \sup_{\tau} \left| Y(\tau) - \mathbb{E}_{\omega'_k} Y'(\tau) \right| && (Y'(\tau) \text{ is zero mean}) \\
 &= \mathbb{E}_{\omega_k} \sup_{\tau} \left| \mathbb{E}_{\omega'_k} [Y(\tau) - Y'(\tau)] \right| && (\text{independence; } \mathbb{E}_{\omega'_k} Y = Y) \\
 &\leq \mathbb{E}_{\omega_k} \mathbb{E}_{\omega'_k} \sup_{\tau} |Y(\tau) - Y'(\tau)| && (\text{Jensen's inequality, } \sup |\cdot| \text{ is convex}) \\
 &= \mathbb{E} \sup_{\tau} |Y(\tau) - Y'(\tau)| && (\text{iterated expectation}) \\
 &= \mathbb{E} \sup_{\tau} |Z(\tau)|.
 \end{aligned}$$

Finally, since $Z(\tau)$ has the same distribution as $Z'(\tau)$, $\mathbb{E} \sup_{\tau} |Z(\tau)| = \mathbb{E} \sup_{\tau} |Z'(\tau)|$.

B Proof of Lemma 10

Recall that, for every $\tau \in T$, $Y(\tau)$ was symmetrized by creating an independent copy $Y'(\tau)$ and defining $Z(\tau) = Y(\tau) - Y'(\tau)$. Therefore, for any $a, \lambda > 0$, the occurrence of the events $\{\sup_{\tau} |Y(\tau)| > a + \lambda\}$ and $\{\sup_{\tau} |Y'(\tau)| < a\}$ imply that

$$\lambda < \sup_{\tau} |Y(\tau)| - \sup_{\tau} |Y'(\tau)| \leq \sup_{\tau} (|Y(\tau)| - |Y'(\tau)|) \leq \sup_{\tau} |Y(\tau) - Y'(\tau)| = \sup_{\tau} |Z(\tau)|.$$

Setting $a = 2\mathbb{E} \sup_{\tau} |Y(\tau)|$ and recalling that $Z(\tau)$ and $Z'(\tau)$ have the same distribution, we conclude that

$$\begin{aligned}
 P \left\{ \sup_{\tau} |Z'(\tau)| > \lambda \right\} &\geq P \left\{ \left\{ \sup_{\tau} |Y(\tau)| > a + \lambda \right\} \cap \left\{ \sup_{\tau} |Y'(\tau)| < a \right\} \right\} \\
 &= P \left\{ \sup_{\tau} |Y(\tau)| > 2\mathbb{E} \sup_{\tau} |Y(\tau)| + \lambda \right\} P \left\{ \sup_{\tau} |Y(\tau)| < 2\mathbb{E} \sup_{\tau} |Y(\tau)| \right\} \\
 &\geq \frac{1}{2} P \left\{ \sup_{\tau} |Y(\tau)| > 2\mathbb{E} \sup_{\tau} |Y(\tau)| + \lambda \right\},
 \end{aligned}$$

where the second line follows from the fact that Y and Y' are independent copies of the same random process, and the third line follows from applying the Markov inequality, which states that for any nonnegative random variable V and any $c > 0$, we have $P\{V \geq c\} \leq c^{-1}\mathbb{E}V$.

C Proof of Lemma 11

Hoeffding's inequality for Rademacher sums [29, Prop. 6.11] states that if b_1, b_2, \dots, b_N are complex numbers and $\epsilon_1, \epsilon_2, \dots, \epsilon_N$ is a Rademacher series, then for every $\lambda \geq 0$ we have

$$P \left\{ \left| \sum_{k=1}^N \epsilon_k b_k \right| > \lambda \right\} \leq 2 \exp \left(\frac{-\lambda^2}{2 \sum_k |b_k|^2} \right). \quad (64)$$

Conditioned on $\{\omega_k\}$ and $\{\omega'_k\}$, we can write

$$Z'(\tau) = \sum_{k=1}^m \epsilon_k \alpha_k, \quad \text{for some } \alpha_k \text{ with } |\alpha_k| = \left| A |\widehat{s}(\omega_k)|^2 e^{i\omega_k(\tau-\tau_0)} - A |\widehat{s}(\omega'_k)|^2 e^{i\omega'_k(\tau-\tau_0)} \right|,$$

and therefore,

$$\begin{aligned} \sum_{k=1}^m |\alpha_k|^2 &\leq |A|^2 \sum_{k=1}^m \left| |\widehat{s}(\omega_k)|^2 e^{i\omega_k(\tau-\tau_0)} - |\widehat{s}(\omega'_k)|^2 e^{i\omega'_k(\tau-\tau_0)} \right|^2 \\ &\leq |A|^2 \sum_{k=1}^m 2 |\widehat{s}(\omega_k)|^4 + 2 |\widehat{s}(\omega'_k)|^4 \\ &= 2|A|^2 M_3^2, \end{aligned}$$

where the second line uses (33). Plugging into (64) yields (35) as desired.

For the increment bound, conditioned on $\{\omega_k\}$ and $\{\omega'_k\}$, we can write

$$Z'(\tau_1) - Z'(\tau_2) = \sum_{k=1}^m \epsilon_k \beta_k,$$

where

$$\begin{aligned} \sum_{k=1}^m |\beta_k|^2 &= |A|^2 \sum_{k=1}^m \left| |\widehat{s}(\omega_k)|^2 e^{i\omega_k(\tau_1-\tau_0)} - |\widehat{s}(\omega'_k)|^2 e^{i\omega'_k(\tau_1-\tau_0)} - |\widehat{s}(\omega_k)|^2 e^{i\omega_k(\tau_2-\tau_0)} + |\widehat{s}(\omega'_k)|^2 e^{i\omega'_k(\tau_2-\tau_0)} \right|^2 \\ &\leq 8|A|^2 \sum_{k=1}^m \left(|\widehat{s}(\omega_k)|^4 \left| \sin \left(\frac{1}{2}\omega_k(\tau_1-\tau_2) \right) \right|^2 + |\widehat{s}(\omega'_k)|^4 \left| \sin \left(\frac{1}{2}\omega'_k(\tau_2-\tau_1) \right) \right|^2 \right) \\ &\leq 8|A|^2 \sum_{k=1}^m |\widehat{s}(\omega_k)|^4 \left| \frac{1}{2}\omega_k(\tau_1-\tau_2) \right|^2 + |\widehat{s}(\omega'_k)|^4 \left| \frac{1}{2}\omega'_k(\tau_2-\tau_1) \right|^2 \\ &\leq 8|A|^2 \sum_{k=1}^m |\widehat{s}(\omega_k)|^4 \left| \frac{|\Omega|}{4}(\tau_1-\tau_2) \right|^2 + |\widehat{s}(\omega'_k)|^4 \left| \frac{|\Omega|}{4}(\tau_2-\tau_1) \right|^2 \\ &\leq \frac{1}{2}|A|^2 M_3^2 |\Omega|^2 |\tau_1-\tau_2|^2. \end{aligned}$$

The second line above follows from applying (33) and then (32), the third line uses the fact that $|\sin(a)| \leq |a|$, and the fourth line follows from the fact that Ω is symmetric about the origin, i.e., that $\Omega = [-\omega_{\max}, \omega_{\max}]$. Plugging into (64) yields (36) as desired.

D Proof of Proposition 13

It follows from (41) that

$$\mathbb{E} \max_i |V_i| = \int_0^\infty \mathbb{P} \left\{ \max_i |V_i| > \lambda \right\} d\lambda.$$

By breaking the integration interval at $\lambda_0 := \sigma\sqrt{2\log(KN)}$, we can write

$$\begin{aligned}
\mathbb{E} \max_i |V_i| &= \int_0^{\lambda_0} \mathbb{P} \left\{ \max_i |V_i| > \lambda \right\} d\lambda + \int_{\lambda_0}^{\infty} \mathbb{P} \left\{ \max_i |V_i| > \lambda \right\} d\lambda \\
&\leq \int_0^{\lambda_0} d\lambda + \int_{\lambda_0}^{\infty} N \max_i \mathbb{P} \{|V_i| > \lambda\} d\lambda \\
&\leq \lambda_0 + \int_{\lambda_0}^{\infty} K N e^{-\lambda^2/2\sigma^2} d\lambda \\
&\leq \lambda_0 + KN \cdot \frac{\sigma^2}{\lambda_0} e^{-\lambda_0^2/2\sigma^2} \\
&= \sigma \left(\sqrt{2\log(KN)} + \frac{1}{\sqrt{2\log(KN)}} \right),
\end{aligned}$$

as claimed. The second line above uses the union bound, and the fourth line uses (34).

E Proof of Lemma 14

The proof essentially follows [22, p. 134]. Fix t_1, t_2, \dots, t_N in Δ and form the vector $g := [G(t_1), G(t_2), \dots, G(t_N)]^T \in \mathbb{C}^N$ with covariance matrix $\Gamma = \Xi \Xi^*$. This vector has the same distribution as Ξh , where $h \in \mathbb{R}^N$ is the standard Gaussian vector, whose entries are i.i.d. zero-mean Gaussian random variables with unit variance. This is because $\mathbb{E}\Xi h = \Xi \cdot \mathbb{E}h = 0_N = \mathbb{E}g$ and $\mathbb{E}\{(\Xi h)(\Xi h)^*\} = \Xi \mathbb{E}\{hh^T\} \Xi^* = \Xi \Xi^* = \Gamma = \mathbb{E}\{gg^*\}$. Here 0_N denotes the $N \times 1$ zero vector. Now consider the function $F(x) : \mathbb{R}^N \rightarrow \mathbb{R}$ defined as $F(x) := \max_{1 \leq i \leq N} |(\Xi x)_i| = \|\Xi x\|_\infty$. Let $x_1, x_2 \in \mathbb{R}^N$. Now we can write

$$\begin{aligned}
|F(x_1) - F(x_2)| &= |\|\Xi x_1\|_\infty - \|\Xi x_2\|_\infty| \\
&\leq \|\Xi(x_1 - x_2)\|_\infty \\
&\leq \|\Xi\|_{\infty,2} \|x_1 - x_2\|_2.
\end{aligned}$$

The second line follows from the triangle inequality, and in the third line, $\|\Xi\|_{\infty,2}$ denotes the operator norm of Ξ from \mathbb{R}^N equipped with the l_2 -norm to \mathbb{C}^N equipped with the l_∞ -norm.

Let B_2^N denote the unit l_2 -ball in \mathbb{R}^N . Note that we have

$$\begin{aligned}
\|\Xi\|_{\infty,2}^2 &= \left(\sup_{x \in B_2^N} \max_{1 \leq i \leq N} |(\Xi x)_i| \right)^2 \\
&= \sup_{x \in B_2^N} \max_{1 \leq i \leq N} |(\Xi x)_i|^2 \\
&= \max_{1 \leq i \leq N} \sup_{x \in B_2^N} \left| \sum_{j=1}^N \Xi_{i,j} x_j \right|^2 \\
&= \max_{1 \leq i \leq N} \sum_{j=1}^N |\Xi_{i,j}|^2 \\
&= \max_i (\Xi \Xi^*)_{i,i} \\
&= \max_i \Gamma_{i,i} \\
&= \max_i \mathbb{E} |G(t_i)|^2 \\
&=: \nu_N^2.
\end{aligned}$$

The third line is a consequence of the Cauchy-Schwartz inequality for complex-valued numbers. Therefore, $F(\cdot)$ is a ν_N -Lipschitz function of its argument (which is a standard Gaussian vector here). So we can invoke, for example, [22, eq. (2.35)] to get

$$\begin{aligned}
\Pr \{F(h) \geq \mathbb{E} F(h) + \lambda\} &= \Pr \left\{ \max_{1 \leq i \leq N} |(\Xi h)_i| \geq \mathbb{E} \max_{1 \leq i \leq N} |(\Xi h)_i| + \lambda \right\} \\
&= \Pr \left\{ \max_{1 \leq i \leq N} |G(t_i)| \geq \mathbb{E} \max_{1 \leq i \leq N} |G(t_i)| + \lambda \right\} \\
&\leq \exp \left(-\frac{\lambda^2}{2\nu_N^2} \right).
\end{aligned}$$

Now we can apply monotone convergence to the above inequality; as $N \rightarrow \infty$, we have $\nu_N^2 \rightarrow \nu^2$ and we come to the desired result.

References

- [1] G. Bennett. Probability inequalities for the sum of independent random variables. *J. Amer. Statist. Assoc.*, 57(297):33–45, 1962.
- [2] F. Beutler. Alias-free randomly timed sampling of stochastic processes. *IEEE Trans. Inform. Theory*, 16(2):147–152, 1970.
- [3] P. Brémaud and A. Ridolfi. Power spectral measure and reconstruction error of randomly sampled signals. In *Proc. IEEE Inform. Theory Workshop*, pages 159–162, 2002.
- [4] E. Candès and T. Tao. Near-optimal signal recovery from random projections: Universal encoding strategies? *IEEE Trans. Inform. Theory*, 52(12):5406–5245, December 2006.

- [5] E. J. Candès, J. Romberg, and T. Tao. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. Inform. Theory*, 52(2):489–509, 2006.
- [6] M. A. Davenport, P. T. Boufounos, M. B. Wakin, and R. G. Baraniuk. Signal Processing with Compressive Measurements. *IEEE J. Sel. Top. Signal Proc.*, 4:445–460, 2010.
- [7] M. A. Davenport, M. F. Duarte, M. B. Wakin, J. N. Laska, D. Takhar, K. F. Kelly, and R. G. Baraniuk. The smashed filter for compressive classification and target recognition. *Proc. Computational Imaging V at SPIE Electronic Imaging*, 2007.
- [8] M. A. Davenport, J. N. Laska, J. R. Treichler, and R. G. Baraniuk. The pros and cons of compressive sensing for wideband signal acquisition: Noise folding vs. dynamic range. *Arxiv preprint arXiv:1104.4842*, 2011.
- [9] D. L. Donoho. Compressed sensing. *IEEE Trans. Inform. Theory*, 52(4):1289–1306, 2006.
- [10] M. F. Duarte and R. G. Baraniuk. Spectral compressive sensing. 2010. Preprint.
- [11] M. F. Duarte, M. A. Davenport, M. B. Wakin, and R. G. Baraniuk. Sparse Signal Detection from Incoherent Projections. In *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Proc. (ICASSP)*, 2006.
- [12] R. M. Dudley. The sizes of compact subsets of Hilbert space and continuity of Gaussian processes. *J. Functional Analysis*, 1(3):290–330, 1967.
- [13] Y. C. Eldar. Compressed sensing of analog signals in shift-invariant spaces. *IEEE Trans. Signal Process.*, 57(8):2986–2997, 2009.
- [14] K. Gedalyahu and Y. C. Eldar. Time-delay estimation from low-rate samples: A union of subspaces approach. *IEEE Trans. Signal Process.*, 58(6):3017–3031, 2010.
- [15] A. C. Gilbert, M. J. Strauss, and J. A. Tropp. A tutorial on fast Fourier sampling. *IEEE Signal Proc. Mag.*, 25(2):57–66, 2008.
- [16] J. Haupt, R. Castro, R. Nowak, G. Fudge, and A. Yeh. Compressive sampling for signal classification. In *Proc. Asilomar Conference on Signals, Systems and Computers*, pages 1430–1434, 2007.
- [17] J. Haupt and R. Nowak. Compressive sampling for signal detection. In *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Proc. (ICASSP)*, Apr. 2007.
- [18] M. A. Jack, P. M. Grant, and J. H. Collins. The theory, design, and applications of surface acoustic wave Fourier-transform processors. *Proceedings of the IEEE*, 68(4):450–468, 1980.
- [19] N. L. Johnson, S. Kotz, and N. Balakrishnan. *Continuous Univariate Distributions*, volume 1. John Wiley and Sons, 2005.
- [20] S. Kunis and H. Rauhut. Random sampling of sparse trigonometric polynomials. *Appl. Comp. Harmonic Analysis*, 22, 2006.

- [21] B. Le, T. W. Rondeau, J. H. Reed, and C. W. Bostian. Analog-to-digital converters: A review of the past, present, and future. *IEEE Signal Proc. Mag.*, 22:69–77, 2005.
- [22] M. Ledoux. *The Concentration of Measure Phenomenon*. American Mathematical Society, 2005.
- [23] M. Ledoux and M. Talagrand. *Probability in Banach Spaces*. Springer, 1991.
- [24] K. S. Lii and E. Masry. Spectral estimation of continuous-time stationary processes from random sampling. *Stochastic Processes and Their Applications*, 52(1):39–64, 1994.
- [25] M. Lustig, D. L. Donoho, J. M. Santos, and J. M. Pauly. Compressed sensing MRI. *IEEE Signal Proc. Mag.*, 25(2):72–82, 2008.
- [26] E. Masry. Poisson sampling and spectral estimation of continuous-time processes. *IEEE Trans. Inform. Theory*, 24(2):173–183, 1978.
- [27] M. W. Meckes. On the spectral norm of a random Toeplitz matrix. *Electronic Communications in Probability*, 12:315–325, 2007.
- [28] H. Rauhut. Circulant and Toeplitz matrices in compressed sensing. In *Proc. SPARS’09*, Saint-Malo, France, 2009.
- [29] H. Rauhut. Compressive sensing and structured random matrices. In *Theoretical Foundations and Numerical Methods for Sparse Recovery*, volume 9 of *Radon Series Comp. Appl. Math.*, pages 1–92. deGruyter, 2010.
- [30] H. Rauhut, J. Romberg, and J. Tropp. Restricted isometries for partial random circulant matrices. *Submitted to Appl. and Comp. Harm. Analysis*, October 2010.
- [31] D. C. Rife and R. Boorstyn. Single tone parameter estimation from discrete-time observations. *IEEE Trans. Inform. Theory*, 20(5):591–598, 1974.
- [32] J. Romberg. A uniform uncertainty principle for Gaussian circulant matrices. In *Proc. Int. Conf. Digital Signal Process.*, 2009.
- [33] J. Romberg and R. Neelamani. Sparse channel separation using random probes. 2010. Preprint.
- [34] M. Rudelson and R. Vershynin. On sparse reconstruction from Fourier and Gaussian measurements. *Comm. Pure Appl. Math.*, 61(8):1025–1045, 2008.
- [35] J.A. Tropp, J.N. Laska, M.F. Duarte, J.K. Romberg, and R.G. Baraniuk. Beyond Nyquist: Efficient Sampling of Sparse Bandlimited Signals. *IEEE Trans. Inform. Theory*, 56(1):520–544, 2010.